HP StorageWorks

Clustered File System 3.0

administration guide

392371–001

Part number: 392371–001
First edition: May 2005

invent

Legal and notice information

Neither PolyServe, Inc. nor Hewlett-Packard Company makes any warranty of any kind with regard to this material, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. Neither PolyServe nor Hewlett-Packard shall be liable for errors contained herein or for incidental or consequential damages in connection with the furnishing, performance, or use of this material.

Clustered File System Administration Guide

Contents

HP Technical Support

1 Introduction

2 Cluster Administration

3 Configure Servers

4 Configure Network Interfaces

5 Configure the SAN

6 Configure Dynamic Volumes

7 Configure PSFS Filesystems

8 Manage Hardware Snapshots

11 Configure Device Monitors

12 Configure Notifiers

13 Test Your Configuration

14 Advanced Topics

15 SAN Maintenance

16 Other Cluster Maintenance

A Management Console Icons

B Error and Log File Messages

Index

HP Technical Support

Telephone numbers for worldwide technical support are listed on the following HP web site: http://www.hp.com/support. From this web site, select the country of origin. For example, the North American technical support number is 800-633-3600.

NOTE: For continuous quality improvement, calls may be recorded or monitored.

Be sure to have the following information available before calling:

- Technical support registration number (if applicable)
- Product serial numbers
- Product model names and numbers
- Applicable error messages
- Operating system type and revision level
- Detailed, specific questions

HP Storage Web Site

The HP web site has the latest information on this product, as well as the latest drivers. Access the storage site at: http://www.hp.com/country/us/eng/prodserv/storage.html. From this web site, select the appropriate product or solution.

HP NAS Services Web Site

The HP NAS Services site allows you to choose from convenient HP Care Pack Services packages or implement a custom support solution delivered by HP ProLiant Storage Server specialists and/or our certified service partners. For more information see us at http://www.hp.com/hps/storage/ns_nas.html.

1

Introduction

HP Clustered File System provides a cluster structure for managing a group of network servers and a Storage Area Network (SAN) as a single entity.

Product Features

HP Clustered File System provides the following features:

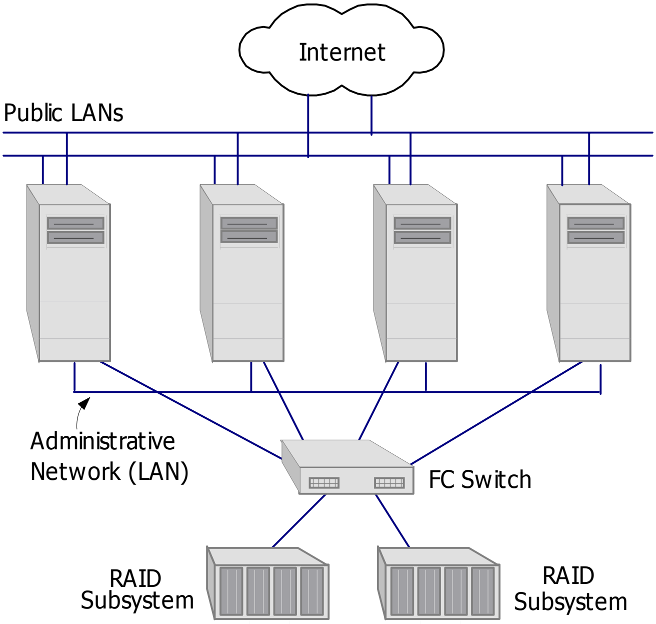
- Fully distributed data-sharing environment. The PSFS filesystem enables all servers in the cluster to directly access shared data stored on a SAN. After a PSFS filesystem has been created on a SAN disk, all servers in the cluster can mount the filesystem and subsequently perform concurrent read and/or write operations to that filesystem. PSFS is a journaling filesystem and provides online crash recovery.
- Availability and reliability. Servers and SAN components (FibreChannel switches and RAID subsystems) can be added to a cluster with minimal impact, as long as the operation is supported by the underlying operating system. HP Clustered File System includes failover mechanisms that enable the cluster to continue operations without interruption when various types of failures occur. If network communications fail between any or all servers in the cluster, HP Clustered File System maintains the coherency and integrity of all shared data in the cluster.

- Cluster-wide administration. The Management Console (a Java-based graphical user interface) and the corresponding command-line interface enable you to configure and manage the entire cluster either remotely or from any server in the cluster.
- Failover support for network applications. HP Clustered File System uses virtual hosts to provide highly available client access to mission-critical data for Web, e-mail, file transfer, and other TCP/IP-based applications. If a problem occurs with a network application, with the network interface used by the virtual host, or with the underlying server, HP Clustered File System automatically switches network traffic to another server to provide continued service.
- Administrative event notification. When certain events occur in the cluster, HP Clustered File System can send information about the events to the system administrator via e-mail, a pager, the Management Console, or another user-defined process.

Overview

The Structure of a Cluster

A cluster includes the following physical components.



Servers. Each server must be running HP Clustered File System.

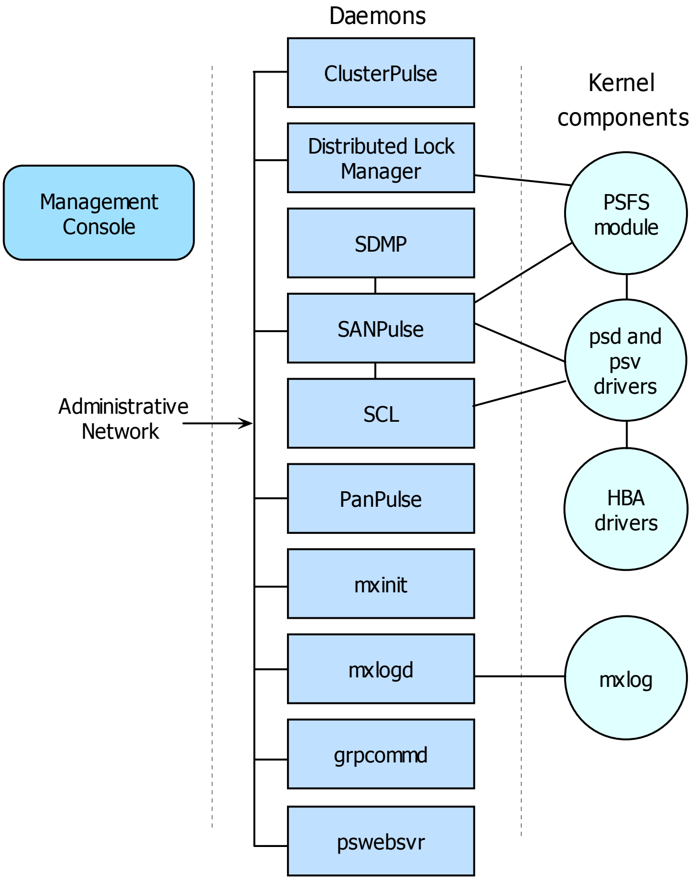
Public LANs. A cluster can include up to four network interfaces per server. Each network interface can be configured to support multiple virtual hosts, which provide failover protection for Web, e-mail, file transfer, and other TCP/IP-based applications.

Administrative Network. HP Clustered File System components communicate with each other over a common LAN. The network used for this traffic is called the administrative network. When you configure the cluster, you can specify the networks that you prefer to use for the administrative network traffic. For performance reasons, we recommend that these networks be isolated from the networks used by external clients to access the cluster.

Storage Area Network (SAN). The SAN includes FibreChannel switches and RAID subsystems. Disks in a RAID subsystem are imported into the cluster and managed from there. After a disk is imported, you can create PSFS filesystems on it.

Software Components

The HP Clustered File System software is installed on each server in the cluster and includes the following major components.



Management Console. Provides a graphical interface for configuring an HP Clustered File System cluster and monitoring its operation. The console can be run either remotely or from any server in the cluster.

ClusterPulse daemon. Monitors the cluster, controls failover of virtual hosts and devices, handles communications with the Management Console, and manages device monitors, service monitors, and event notification.

Distributed Lock Manager (DLM) daemon. Provides a locking mechanism to coordinate server access to shared resources in the cluster. All reads and writes to a PSFS filesystem automatically obtain the appropriate locks from the DLM, ensuring filesystem coherency.

SDMP daemon. Arbitrates and ensures that only one cluster has access to shared SAN devices, in the event that a server loses connectivity.

SANPulse daemon. Provides the cluster infrastructure for management of the SAN. SANPulse coordinates filesystem mounts, unmounts, and crash recovery operations.

PSFS filesystem module. The shared filesystem.

SCL daemon. Manages shared storage devices. The SCL has two primary responsibilities: assigning device names to shared disks when they are imported into the cluster, and enabling or disabling a server's access to shared storage devices.

psd driver. Provides cluster-wide consistent device names among all servers.

psv driver. Used by the Clustered Volume Manager, which creates, extends, recreates, or destroys dynamic volumes.

HBA drivers. HP Clustered File System includes drivers for the supported FibreChannel host bus adapters.

PanPulse daemon. Selects and monitors the network to be used for the administrative network, verifies that all hosts in the cluster can communicate with each other, and detects any communication problems.

mxinit daemon. Starts or stops HP Clustered File System and monitors HP Clustered File System processes.

mxlogd daemon. Manages global error and event messages. The messages are written to the */var/log/hpcfs/matrix.log* file on each server.

mxlog module. Allows HP Clustered File System kernel modules to send messages to the mxlogd daemon.

grpcommd daemon. Manages HP Clustered File System group communications across the cluster.

pswebsvr daemon. The embedded web server daemon used by the Management Console and the **mx** utility.

Administrative Network. Handles HP Clustered File System administrative traffic. Most HP Clustered File System daemons communicate with each other over the administrative network.

Shared SAN Devices

Before a SAN disk can be used, you will need to import it into the cluster. This step gives HP Clustered File System complete and exclusive control over access to the disk. During the import, the disk is given a unique global device name. The servers in the cluster use this name when they need to access the disk.

PSFS Filesystems

PSFS filesystems can be created on either basic volumes or dynamic volumes. A basic volume is a single disk, disk partition, or LUN imported into the cluster. Dynamic volumes consist of one or more imported disks, disk partitions, or LUNs and are created by the Clustered Volume Manager.

The PSFS filesystem provides the following features:

- Concurrent access by multiple servers. After a filesystem has been created on a shared disk, all servers having physical access to the device via the SAN can mount the filesystem. A PSFS filesystem must be consistently mounted either read-only or read-write across the cluster.

- Support for standard filesystem operations such as **mkfs**, **mount**, and **umount**. These operations can be performed either with the Management Console or from the command line.

- Support for existing applications. The PSFS filesystem uses standard read/write semantics and does not require changes to applications.
- Journaling and live crash recovery. Filesystem metadata operations are written to a journal before they are performed. If a server using the filesystem should crash during an operation, the journal is replayed and any journaled operations in progress at the time of the crash are completed. Users on other servers will experience only a slight delay in filesystem operations during the recovery.

HP Clustered File System Databases

HP Clustered File System uses the following databases to store cluster information:

- **Shared Memory Data Store (SMDS).** The SANPulse daemon stores filesystem status information in this database. The database contains *cp_status* and *sp_status* files that are located in the directory */var/hpcfs/run* on each server. These files should not be changed.
- **Device database.** The SCL assigns a device name to each shared disk imported into the cluster. It then stores the device name and the physical UID of the disk in the device database. The database is located on the membership partitions that you selected when installing the HP Clustered File System product. (The membership partitions are also used for functions related to SAN control.)
- **Volume database.** This database stores information about dynamic volumes and is located on the membership partitions.

Virtual Hosts and Failover Protection

HP Clustered File System uses virtual hosts to provide failover protection for servers and network applications.

A *virtual host* is a hostname/IP address configured on one or more servers. The network interfaces selected on those servers to participate in the virtual host must be on the same subnet. One server is selected as the primary server for the virtual host. The remaining servers are backups.

The primary and backup servers do not need to be dedicated to these activities; all servers can support other independent functions.

To ensure the availability of a virtual host, HP Clustered File System monitors the health of all network interfaces and the health of the underlying server. If you have created service or device monitors, those monitors periodically check the health of the specified services or devices. If any of these health checks fail, HP Clustered File System can transfer the virtual host to a backup server and the network traffic will continue.

After creating virtual hosts, you will need to configure your network applications to recognize them. When clients want to access a network application, they use the virtual host address instead of the address of the server where the application is running.

Service and Device Monitors

A *service* is a network service such as HTTP or FTP that is installed and configured on the servers in the cluster. HP Clustered File System can be configured to watch specific services with *service monitors*.

A service monitor is created on a virtual host. The service being monitored should be installed on all servers associated with that virtual host. When a service monitor determines that a service has failed, HP Clustered File System transfers the virtual host for that service to a backup server. For example, if the HTTP service fails, HP Clustered File System will transfer the HTTP traffic to a backup server that provides HTTP.

HP Clustered File System includes several built-in service monitors for monitoring well-known network services. You can also configure custom monitors for other services.

A *device monitor* is similar to a service monitor; however, it is designed either to watch a part of a server such as a local disk drive or to monitor a PSFS filesystem or a gateway device. A device monitor is assigned to one or more servers. HP Clustered File System provides several built-in device monitors. You can also define your own custom device monitors.

By default, when a device monitor is assigned to a server, all virtual hosts on that server are dependent on the device monitor. However, you can select the virtual hosts that will be dependent on the device monitor. If a device monitor indicates that a device is not functioning properly on the primary server, HP Clustered File System transfers the dependent virtual host addresses from the primary server to a backup server.

Event Notification

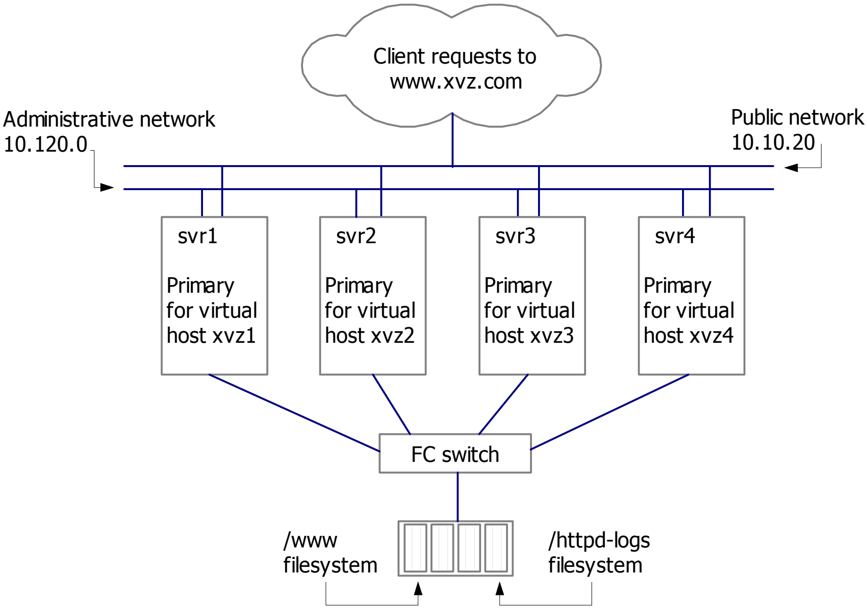
A *notifier* defines how the cluster handles state transition, error, warning, and informational messages. You can configure a notifier with a combination of events and originating cluster entities and then supply a script that specifies the action the notifier should take. For example, you could forward events to e-mail or to any other user-defined process.

A Sample Configuration

This sample configuration is for a cluster providing a Web site. HP Clustered File System ensures high availability of the Web data for incoming client requests and also guarantees that the site administrator can access the Web logs and update the Web data as needed.

The cluster includes four servers and two networks. The public network, 10.10.20, carries the traffic for client requests. The private network, 10.120.0, is used as the administrative network. (The administrative network will fail over to the public network temporarily if the private network should become unavailable.)

DNS round robin is used to rotate the incoming client requests among the four servers in the cluster. Clients access the web site via the address www.xvz.com. These requests are routed in turn to four virtual hosts that are associated with the four servers in the cluster. Each server is primary for one of the virtual hosts and backup for the other virtual hosts.


Client requests to www.xvz.com
Administrative network 10.120.0
Public network 10.10.20
svr1 Primary for virtual host xvz1
svr2 Primary for virtual host xvz2
svr3 Primary for virtual host xvz3
svr4 Primary for virtual host xvz4
FC switch
/www filesystem
/httpd-logs filesystem

Two PSFS filesystems are mounted on each server. The first filesystem contains the web data accessed via the virtual hosts and is mounted at */www* on each server. The directory */var/www/html* is a symbolic link to this mountpoint. The second filesystem contains the log files generated by the Web activity and is mounted at */httpd-logs* on each server. The directory */var/log/httpd* is a symbolic link to this mountpoint.

Failover Support for Client Requests

The configuration includes virtual hosts, HTTP service monitors, and SHARED_FILESYSTEM device monitors to ensure that clients can always access the Web site.

- If a server (or its network interface) fails, any virtual hosts that are active on that server (or network interface) will fail over to a backup server. The client requests will then be processed on that server.
- An HTTP service monitor is configured on each server. If this monitor detects that the HTTP service has failed, the associated virtual host

will fail over to a backup server where the HTTP service is healthy. The client requests will then be processed on that server.

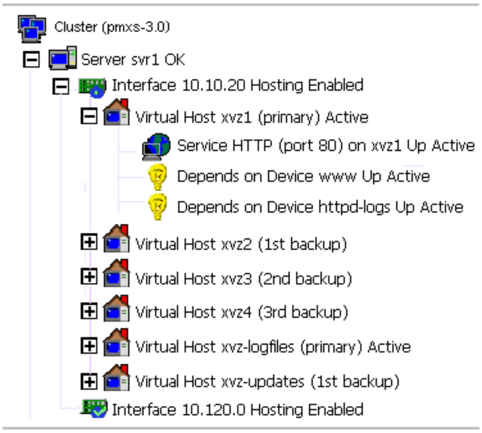
- Two SHARED_FILESYSTEM device monitors are configured on each server for the PSFS filesystems (*/www* and */httpd-logs*). These monitors check the mount status and health of the mounted PSFS filesystems. The four virtual hosts are also associated with each monitor. If a monitor detects a problem with a filesystem, any active virtual hosts associated with the monitor will fail over to a backup server where the filesystem mount is healthy. The client requests will then be processed on that server.

Support for Administrative Tasks

Because the Web site data and access logs are on shared PSFS filesystems, they can be accessed from any server in the cluster. A server failure does not prevent the administrator from accessing log files or updating the Web site data.

The sample configuration includes two virtual hosts, xvz-logfiles and xvz-updates, that are used to run custom scripts that process log file data or perform the updates. The use of virtual hosts ensures that the administrator will always reach an "up" server when it is necessary to perform administrative tasks.

The complete configuration for svr1 looks like this.



```
Cluster (pmxs-3.0)
  ⊟ Server svr1 OK
      ⊟ Interface 10.10.20 Hosting Enabled
          ⊟ Virtual Host xvz1 (primary) Active
                Service HTTP (port 80) on xvz1 Up Active
                Depends on Device www Up Active
                Depends on Device httpd-logs Up Active
          ⊞ Virtual Host xvz2 (1st backup)
          ⊞ Virtual Host xvz3 (2nd backup)
          ⊞ Virtual Host xvz4 (3rd backup)
          ⊞ Virtual Host xvz-logfiles (primary) Active
          ⊞ Virtual Host xvz-updates (1st backup)
        Interface 10.120.0 Hosting Enabled
```

Cluster Design Guidelines

Be sure to consider the following guidelines when planning the physical configuration of your HP Clustered File System cluster.

Server Memory

Memory resources are consumed on each cluster server to manage the state necessary to preserve the coherency of shared filesystems. For this reason, the servers in the cluster should have approximately equal amounts of physical memory. As a guideline, the ratio of memory between the smallest and largest servers should not exceed 2. For example, the smallest server could have 1 GB of memory and the largest server 2 GB. If this ratio is exceeded (such as a 256 MB server and a 2 GB server, with a ratio of 8), paging can increase on the smallest server to the extent that overall cluster performance is significantly reduced.

Multipath I/O

Multipath I/O can be used in a cluster configuration to eliminate single points of failure. It supports the following:

- Up to four FibreChannel ports per server. If an FC port or its connection to the fabric should fail, the server can use another FC port to reach the fabric.

- Multiple FibreChannel switches. When the configuration includes more than one FC switch, the cluster can survive the loss of a switch without disconnecting servers from the SAN. Servers connected to the failed switch can still access the fabric if they have another FC port attached to another FC switch.

- Multiported SAN disks. Multiported disks can have connections to each FC switch in the cluster. If a switch fails, disk access continues uninterrupted if a path exists from a remaining switch.

When you start HP Clustered File System, it automatically configures all paths from each server to the storage devices. On each server, it then uses the first path it discovered for I/O with the SAN devices. If that path fails, HP Clustered File System automatically fails over the I/O to another path.

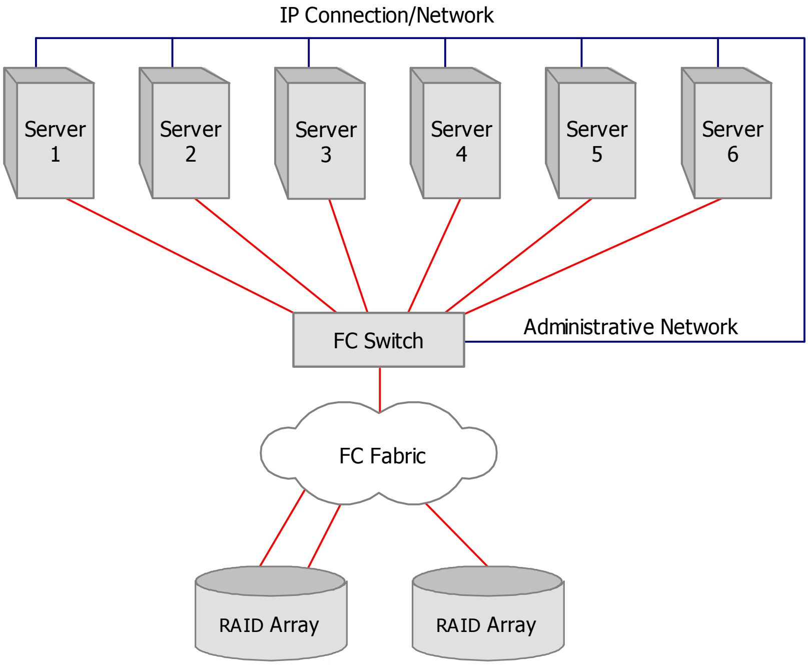
The **mxmpio** command can be used to monitor or manipulate multipath I/O (MPIO) devices. See “Manage Multipath I/O” on page 60 for details.

Supported Configurations

HP Clustered File System supports up to four FibreChannel ports per server, multiple FC switches, and multiported SAN disks. The following diagrams show some sample cluster configurations using these components. In the diagrams, the FC fabric cloud can include additional FC switches that are not managed by HP Clustered File System.

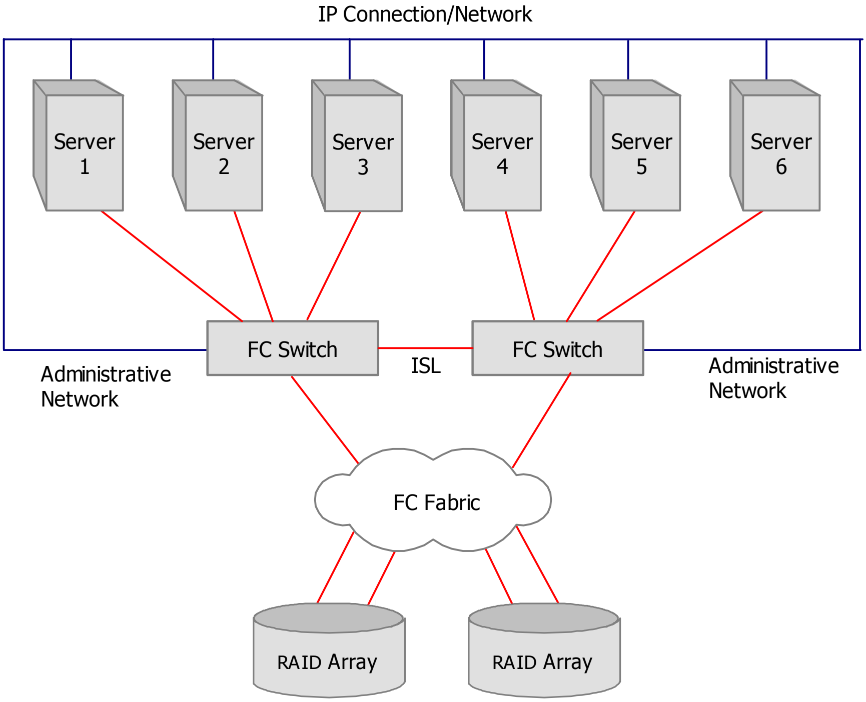
Single FC Port, Single FC Switch, Single FC Fabric

This is the simplest configuration. Each server has a single FC port connected to an FC switch managed by the cluster. The SAN includes two RAID arrays. In this configuration, multiported SAN disks can protect against a port failure, but not a switch failure.



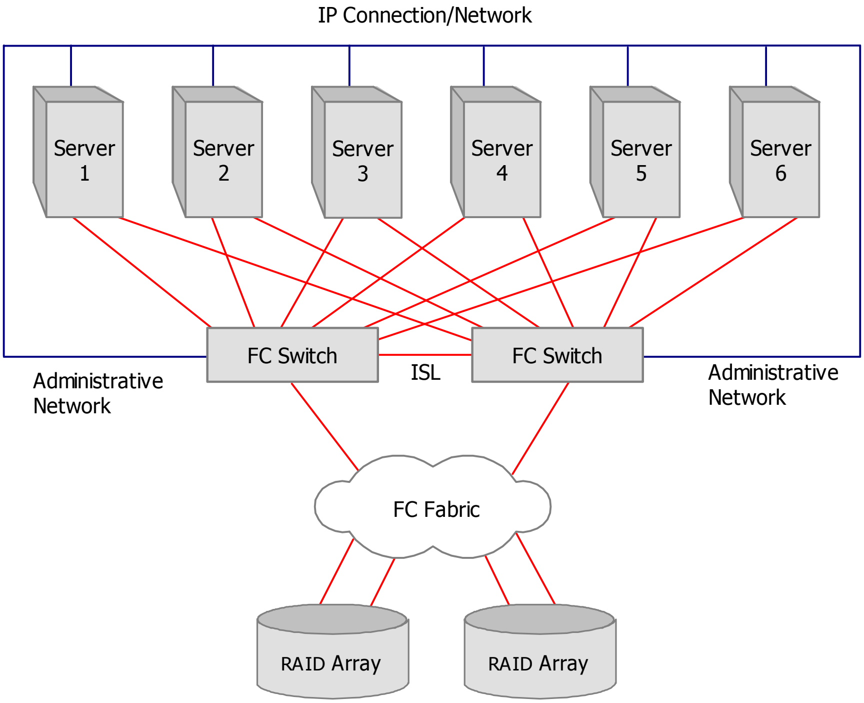
Single FC Port, Dual FC Switches, Single FC Fabric

In this example, the fabric includes two FC switches managed by the cluster. Servers 1–3 are connected to the first FC switch; servers 4–6 are connected to the second switch. The cluster also includes two RAID arrays, which contain multiported disks. If a managed FC switch fails, the servers connected to the other switch will survive and access to storage will be maintained.



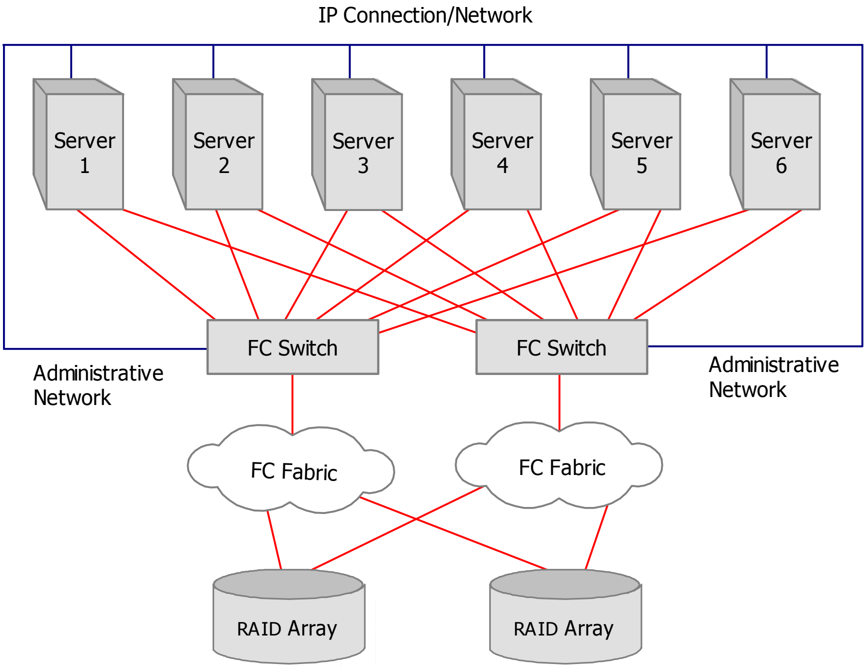
Dual FC Ports, Dual FC Switches, Single FC Fabric

This example uses multipath I/O to eliminate single points of failure. The fabric includes two FC switches managed by the cluster. Each server has two FC ports; the first FC port connects to the first FC switch and the second FC port connects to the second FC switch. The cluster also includes two RAID arrays containing multiported disks.



Dual FC Ports, Dual FC Switches, Dual FC Fabrics

This example is similar to the previous example, but also includes dual FC fabrics, with a cluster-managed FC switch in each fabric. If one of the fabrics should fail, the servers can access the storage via the other fabric.



2

Cluster Administration

HP StorageWorks Clustered File System can be administered either with the Management Console or from the command line.

Administrative Considerations

You should be aware of the following when managing HP Clustered File System:

- Normal operation of the cluster depends on a reliable network hostname resolution service. If the hostname lookup facility becomes unreliable, this can cause reliability problems for the running cluster. Therefore, you should ensure that your hostname lookup services are configured to provide highly reliable lookups, particularly for the hostnames that are critical to cluster operation. These critical hostnames include the hostnames of all servers in the cluster and the hostnames of all FibreChannel switches being used with the cluster.

  One easy way to ensure reliable hostname lookups is to put all critical hostnames and their IP addresses in */etc/hosts* on each of the servers (see the **hosts**(5) man page). After you have done this, ensure that the host lookup order in */etc/nsswitch.conf* specifies that the *hosts* file will be examined first (see the **nsswitch.conf**(5) man page). Other mechanisms, apart from */etc/hosts,* can also be used, depending on your site network configuration.

- For best performance, we recommend that you monitor the cluster from a separate administrative station rather than from a server in the cluster. The Management Console can be installed on Linux or Windows systems.
- HP Clustered File System components can fail if the */var* filesystem fills up. To avoid this situation, move */var/hpcfs* to a separate partition containing at least 200 MB.
- To ensure that file times are consistent across the cluster, it is important that all cluster servers operate with synchronized time-of-day clocks. An NTP server is one commonly used mechanism for synchronizing system clocks.
- When fabric fencing is configured, do not enable or disable the FibreChannel ports used by the cluster. HP Clustered File System must have complete control over these ports to protect the integrity of the shared filesystems on the SAN. HP Clustered File System will reenable a port if it detects that the port has been disabled manually.
- You cannot add or delete a network interface device from the operating system while HP Clustered File System is running. Attempting to do this can adversely affect the entire cluster. If you need to add or delete a network interface on a particular server, first stop HP Clustered File System on that server. Then make your administrative change and restart HP Clustered File System.
- A network interface cannot be modified or taken down while HP Clustered File System is running. If any part of a network interface definition needs to be changed, stop HP Clustered File System on the affected server, shut down the network, and then make the change. When the change is complete, bring up the network and restart HP Clustered File System.
- The Linux kernel can access only partitions 01 through 15 on SCSI disks. However, the **fdisk** and **cfdisk** utilities allow you to create partitions with numbers greater than 15. If you use these utilities for partitioning SCSI disks, be sure that partition 15 is the highest-numbered partition.

- HP Clustered File System supports 2 TB filesystems on basic volumes; however, Linux has a 1 TB partition size limit. This limit prevents the creation of a 2 TB filesystem without modifications to the Linux kernel. HP Clustered File System supports up to 16 TB filesystems on dynamic volumes.
- If servers from multiple clusters can access the SAN via a shared FC fabric, avoid importing the same disk into more than one cluster. Filesystem corruption can occur when different clusters attempt to share the same filesystem.

Tested Configuration Limits

HP Clustered File System configurations have been tested up to the following limits:

- 16 servers per cluster
- 508 LUNs per cluster on SuSE
- 512 filesystems per cluster
- 1750 filesystem mounts per cluster
- 64 virtual hosts per cluster
- 64 service and/or device monitors per cluster (the total number of service and device monitors cannot exceed 64)
- 10 event notifiers per cluster
- 4 network interface cards per server
- 4 FibreChannel ports per server

These limits will be increased as additional testing takes place.

Theoretically, the tested configuration limits can be exceeded up to the bounds of the operating system. If you plan to configure your cluster beyond the tested limits, please contact HP Technical Support for information about any known configuration issues or concerns.

Cluster Management Applications

HP Clustered File System provides two management applications: **mxconsole**, the Management Console, and **mx**, the corresponding command-line interface. These applications can be run from either a Cluster server or a local machine outside the cluster.

On a local machine, the management applications download their rendering and application logic from the servers in the cluster. To minimize startup time, the downloaded software is cached on the local machine. The local machine then uses the software from the cache whenever possible. If for some reason the software version running on the server cannot be identified, the applications use the latest version in their cache.

When you invoke **mxconsole** or **mx** from a local machine, by default the application checks the current software version on the server to which it is being connected and then downloads the software only if that version is not already in the local cache.

Manage a Cluster with the Management Console

The Management Console can manipulate all entities in the cluster, including entities that are currently down. For example, you can change the configuration of a down server; it will learn its new configuration when it comes back up and rejoins the cluster.

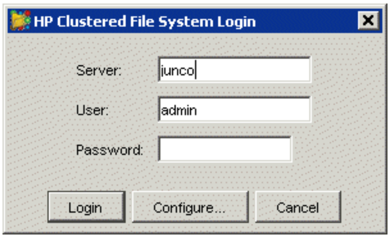
NOTE: For improved performance, the Management Console caches hostname lookups. If your DNS changes, you may need to restart the console so that it will reflect the new hostname.

Start the Management Console

To start the Management Console, first start the windowing environment and then type the following command:

```
$ mxconsole
```

The Login window then appears. If the window does not display properly, verify that your DISPLAY variable is set correctly.



Server: junco

User: admin

Password:

Server: Type the name or IP address of a server in the cluster. If you are creating a new cluster, type the name or IP address of a server that will be in the new cluster.

User: Type the name of a user configured on the server. The administrator is user *admin*. Other users have read permission only.

Password: Type the appropriate password.

Disconnect from a Cluster

To close the Management Console window for the current server, either select File > Disconnect or click the Disconnect icon on the toolbar. You can then use either File > Connect or the Connect icon on the toolbar to connect to a cluster from another server.

Exit a Management Console Session

To end an HP Clustered File System console session, select File > Exit. The cluster will continue to operate after you disconnect from it. You do not need to have an open Management Console connection to a cluster.

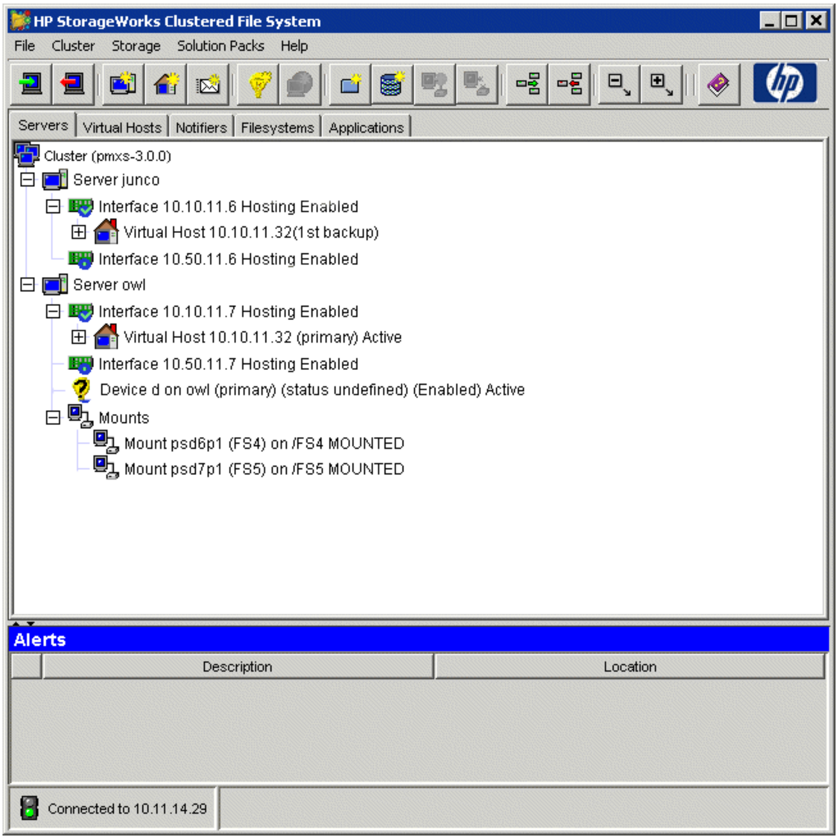
Manage a Cluster from the Command Line

The **mx** utility allows you to manage HP Clustered File System from the command line. See the *HP Clustered File System Command Reference Guide* for more information about this utility.

PSFS filesystems can also be managed with Linux shell commands. Changes made with these commands are reflected on the Management Console.

The Management Console

When you connect to the cluster via **mxconsole**, the Management Console appears. The Management Console window has five tabs that show different views of the cluster. Following is the Servers view.



```
Cluster (pmxs-3.0.0)
  Server junco
    Interface 10.10.11.6 Hosting Enabled
      Virtual Host 10.10.11.32(1st backup)
    Interface 10.50.11.6 Hosting Enabled
  Server owl
    Interface 10.10.11.7 Hosting Enabled
      Virtual Host 10.10.11.32 (primary) Active
    Interface 10.50.11.7 Hosting Enabled
    Device d on owl (primary) (status undefined) (Enabled) Active
    Mounts
      Mount psd6p1 (FS4) on /FS4 MOUNTED
      Mount psd7p1 (FS5) on /FS5 MOUNTED
```

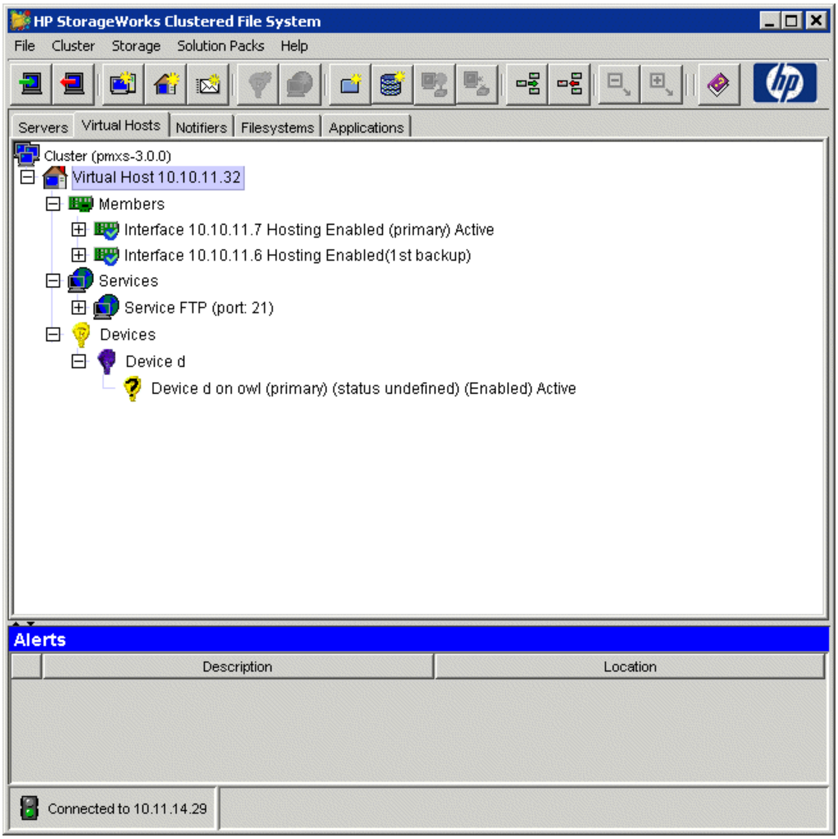
The toolbar at the top of the window can be used to connect or disconnect from a cluster, to add new cluster entities (servers, virtual hosts, notifiers, device monitors, service monitors, and filesystems), to mount or unmount filesystems, to import or deport disks, to collapse or expand the entity lists, and to display online help. “Management Console Icons” on page 215 describes the icons used to represent cluster entries and their status.

Servers Tab

The Servers tab shown above lists the entire configuration of each server configured in the cluster, including the network interfaces on the server, any virtual hosts associated with those interfaces, any device monitors created on the server, and any PSFS filesystems mounted on the server.

Virtual Hosts Tab

The Virtual Hosts tab shows all virtual hosts in the cluster. For each virtual host, the window lists the network interfaces on which the virtual host is configured, any service monitors configured on that virtual host, and any device monitors associated with that virtual host.

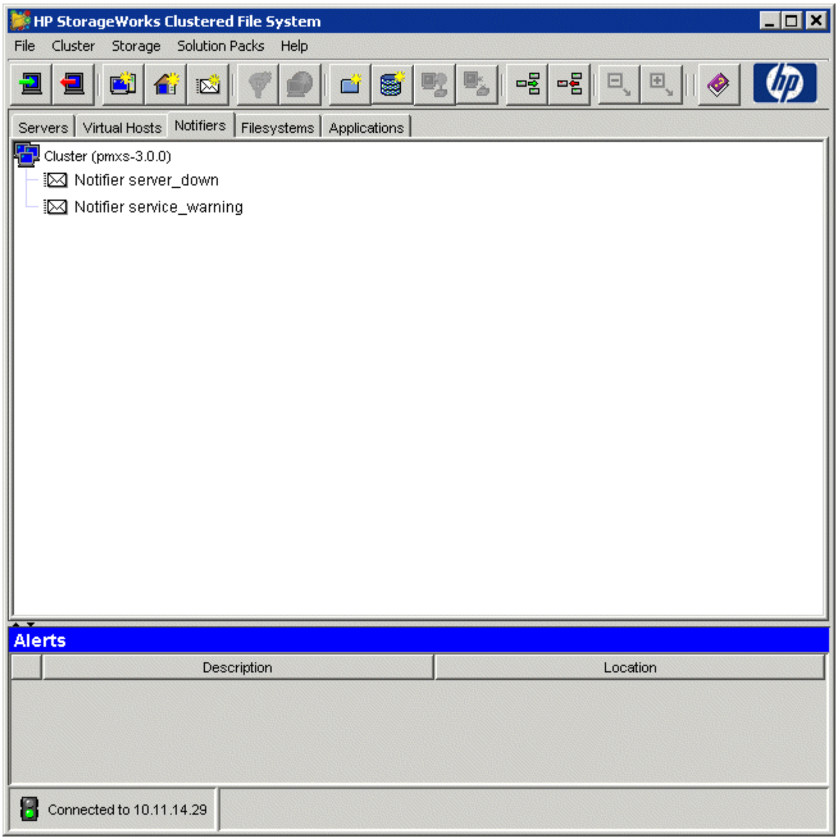


```
Cluster (pmxs-3.0.0)
  Virtual Host 10.10.11.32
    Members
      Interface 10.10.11.7 Hosting Enabled (primary) Active
      Interface 10.10.11.6 Hosting Enabled(1st backup)
    Services
      Service FTP (port: 21)
    Devices
      Device d
        Device d on owl (primary) (status undefined) (Enabled) Active
```

Connected to 10.11.14.29

Notifiers Tab

The Notifiers tab shows all notifiers configured in the cluster.



HP StorageWorks Clustered File System
File Cluster Storage Solution Packs Help
Servers Virtual Hosts Notifiers Filesystems Applications
Alerts
Description
Location

Cluster (pmxs-3.0.0)

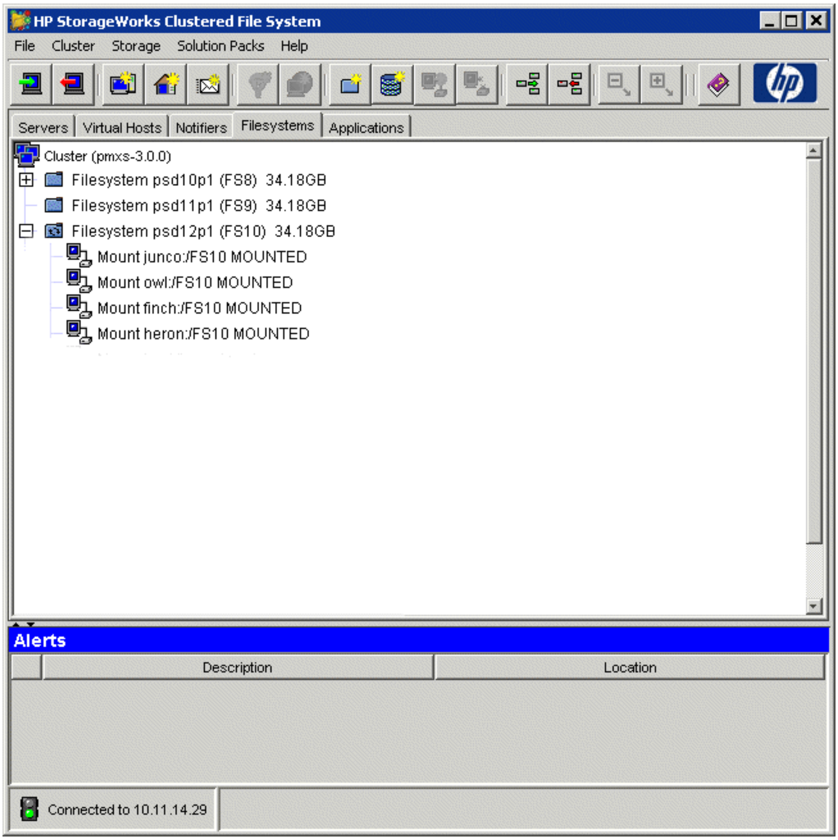
Notifier server_down

Notifier service_warning

Connected to 10.11.14.29

Filesystems Tab

The Filesystems tab shows all PSFS filesystems in the cluster.

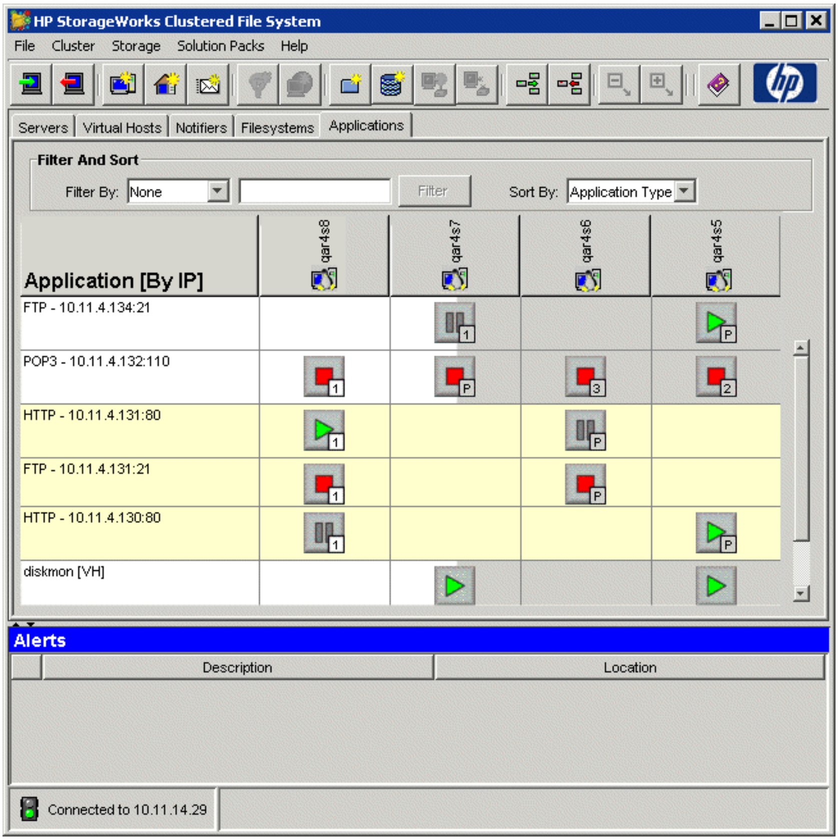


Cluster (pmxs-3.0.0)
Filesystem psd10p1 (FS8) 34.18GB
Filesystem psd11p1 (FS9) 34.18GB
Filesystem psd12p1 (FS10) 34.18GB
Mount junco:/FS10 MOUNTED
Mount owl:/FS10 MOUNTED
Mount finch:/FS10 MOUNTED
Mount heron:/FS10 MOUNTED

Connected to 10.11.14.29

Applications Tab

This view shows the application monitors configured in the cluster and provides the ability to manage and monitor them from a single screen. The tab uses a table format, with a column for each server in the cluster. The application monitors appear in the rows of the table.



Filter By: None

Sort By: Application Type

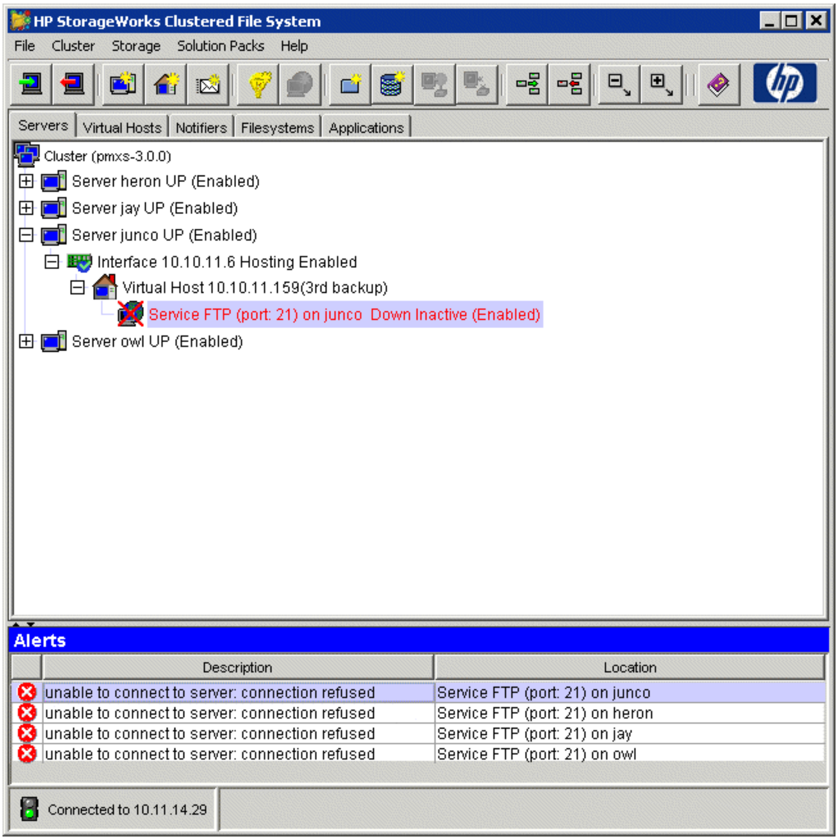
Application [By IP]	qar4s8	qar4s7	qar4s6	qar4s5
FTP - 10.11.4.134:21		⏸ 1		▶ P
POP3 - 10.11.4.132:110	■ 1	■ P	■ 3	■ 2
HTTP - 10.11.4.131:80	▶ 1		⏸ P	
FTP - 10.11.4.131:21	■ 1		■ P	
HTTP - 10.11.4.130:80	⏸ 1			▶ P
diskmon [VH]		▶		▶

Connected to 10.11.14.29

You can reorder the information on this tab or limit the information that is displayed. The Applications tab also allows you to modify the configuration of a virtual host or to move a virtual host to another server. You can also enable, disable, or delete a monitor, view or clear monitor errors, and modify the configuration of a monitor. For more information, see “Using the Applications Tab” on page 111.

Cluster Alerts

The Alerts section at the bottom of the Management Console window lists errors that have occurred in cluster operations. Double click an alert to view the error in the cluster tree structure.



Cluster (pmxs-3.0.0)
- Server heron UP (Enabled)
- Server jay UP (Enabled)
- Server junco UP (Enabled)
  - Interface 10.10.11.6 Hosting Enabled
    - Virtual Host 10.10.11.159(3rd backup)
      - Service FTP (port: 21) on junco Down Inactive (Enabled)
- Server owl UP (Enabled)

	Description	Location
⊗	unable to connect to server: connection refused	Service FTP (port: 21) on junco
⊗	unable to connect to server: connection refused	Service FTP (port: 21) on heron
⊗	unable to connect to server: connection refused	Service FTP (port: 21) on jay
⊗	unable to connect to server: connection refused	Service FTP (port: 21) on owl

Connected to 10.11.14.29

Assign or Change Passwords

If you need to assign a new password or to change an existing password on a particular server, use one of these methods:

Cluster Configuration window. Select File > Configure. You can then change the *admin* password on the General Settings tab.

If the cluster is running, you will need to change the password individually on each server. If you stop HP Clustered File System on all of the servers, you can change the password on one server and then export the new configuration (on the Cluster Wide Configuration tab) to the other servers.

NOTE: You can use the same methods to change the secret network key.

mxpasswd Command. To add a new user and password, use the following syntax, where *user* and *password* are enclosed in curly braces. You must be user *root*.

```
# mxpasswd
mxpasswd> {user} {password}
```

To change an existing password, use the following syntax. Users can change their own passwords. If you are *root*, you can change any user password without specifying the old password.

```
$ mxpasswd
mxpasswd> {user} {new_password} {old_password}
```

HP Clustered File System Processes and mxinit

When HP Clustered File System is running on a server, the following HP Clustered File System processes should be active:

clusterpulse	The ClusterPulse daemon
panpulse	The PanPulse daemon
sanpulse	The SANPulse daemon
dlm	The Distributed Lock Manager daemon
grpcommd	The cluster-wide communications daemon
mxlogd	The log daemon
mxinit	The **mxinit** daemon
pswebsvr	The embedded web server daemon used by the Management Console and the **mx** utility

Do not terminate any of these processes; they are required for HP Clustered File System operations.

Process Monitoring

The **mxinit** utility is started automatically as a daemon on each server and monitors all HP Clustered File System processes running there. (You can start another instance of **mxinit** to perform other tasks provided by the utility.)

If an HP Clustered File System process dies unexpectedly on a server, **mxinit** will restart it if possible. However, certain process failures may force a restart of that particular server.

NOTE: When you invoke **mxinit** to start HP Clustered File System, by default it continues running and monitors processes. If you do not want **mxinit** to monitor processes, invoke it with the **-M** (or **--no-monitor**) option. It will then exit after it completes the options you specified.

View HP Clustered File System Process Status

You can use the **mxinit** utility or the **pmxs** script to display the status of HP Clustered File System processes and modules. To display status information with **mxinit**, either invoke it with no options or specify the **--status** option. You can also use the command **/etc/init.d/pmxs status** to display this information. You do not need to be user *root*.

```
$ /opt/hpcfs/bin/mxinit --status
HP Clustered File System Status:
Name                pid / status
mxinit              15930 mxinit: Monitoring MxS processes
mxlog               Loaded
grpcommd            15885
mxlogd              15887
pswebsvr            15890
clusterpulse        15892
panpulse            15912
psd_devfs           Loaded
psfs                Loaded
dlm                 15913
sanpulse            15917
```

```
FibreChannel adapter module status:
qla2300 - QLogic 2300 FibreChannel Adapter, is Loaded
```

The PID is displayed for running processes; “Stopped” is displayed for processes that are not running. For modules, the status specifies whether the module is loaded. “FibreChannel adapter module status” displays status for the FibreChannel adapter modules installed on the system.

mxinit Configuration File

mxinit performs its actions according to a set of default values. You can use the */etc/hpcfs/mxinit.conf* configuration file to override these values. The file describes the available options and the required format. We recommend that you change this file only under the direction of HP personnel.

Start or Stop HP StorageWorks Clustered File System

HP Clustered File System runs on each server in the cluster. When a server is booted to run-levels 3 or 5, HP Clustered File System is started automatically by the script **/etc/init.d/pmxs**.

Start or Stop HP Clustered File System with the pmxs Script

If you need to start, stop, or restart HP Clustered File System on a particular server, invoke the **pmxs** script:

```
# /etc/init.d/pmxs start
# /etc/init.d/pmxs stop
# /etc/init.d/pmxs restart
```

The **pmxs** script calls the **mxinit** utility. For a start operation, **mxinit** starts the HP Clustered File System processes in the correct order and loads the kernel modules. For a stop operation, it stops the HP Clustered File System processes, unloads the kernel modules, and performs cleanup tasks.

Start or Stop HP Clustered File System with mxinit

Typically, you should use the **pmxs** script to start or stop HP Clustered File System. However, if you want to see verbose output during the start or stop operation, you can run **mxinit** manually with the **--verbose** option.

You can use the following **mxinit** options to start or stop HP Clustered File System:

- **-s**, **--start**

  Start the HP Clustered File System processes.

- **-x**, **--stop**

  Gently stop the HP Clustered File System processes. **mxinit** first attempts to unmount PSFS filesystems. If the unmount fails, the gentle stop operation will also fail.

  If you specify both **--stop** and **--hard**, the **mxinit** command first attempts the **--stop** operation. If it fails, **mxinit** then executes the **--hard** operation.

- **-H**, **--hard**

  Perform a hard, immediate stop of the HP Clustered File System processes. **mxinit** first attempts to terminate any applications accessing PSFS filesystems. It then unmounts the filesystems, terminates the HP Clustered File System processes, and unloads HP Clustered File System modules.

- **-L**, **--load-mod**

  Load all HP Clustered File System modules.

- **-U**, **--unload-mod**

  Unload all HP Clustered File System modules.

- **-v**, **--verbose**

  Print verbose output about each step of the **mxinit** operation.

- **-m**, **--monitor**

  Explicitly tell **mxinit** to monitor processes. This is the default when **mxinit** is invoked to start HP Clustered File System.

- **-M**, **--no-monitor**

  Explicitly tell **mxinit** not to monitor processes.

- **--hba-status**

  Display the state of the FibreChannel host bus adapter drivers.

Back Up and Restore HP Clustered File System

The following HP Clustered File System configuration files should be backed up periodically.

Files in */etc/hpcfs*:

- cp_conf
- fence.conf
- licenses/license
- mxinit.conf
- mx_passwd
- mxsecret
- oem.conf
- psSan.cfg
- scl.conf
- snap.conf

Files in */var/hpcfs*:

- FCswitches
- MPdata

You can use these files to restore the HP Clustered File System configuration if it should be destroyed by accident or by an event such as a hardware failure.

HP Clustered File System Man Pages

HP Clustered File System includes man pages for the commands and utilities provided with the product. To enable the **man** command to locate these man pages, prepend */opt/hpcfs/man* to your MANPATH. The following example is for the Bash shell:

```
export MANPATH=/opt/hpcfs/man:$MANPATH
```

The man pages are as follows:

device(8)	**mx** command to manipulate device monitors
disk(8)	**mx** command to manipulate SAN disks
dynvolume(8)	**mx** command to manage dynamic volumes
exportgroup(8)	**mx** command to manage export groups
fs(8)	**mx** command to manipulate PSFS filesystems
matrix(8)	**mx** command to manipulate a cluster
matrixrc(8)	The cluster configuration file
mkpsfs(8)	Create a PSFS filesystem
mount_psfs(8)	Mount a PSFS filesystem
mx(8)	Manipulate a cluster
mxinit(8)	Start, stop, or monitor HP Clustered File System processes
mxlogger(8)	Add a log message to the cluster log
mxmpconf(8)	Manage membership partitions
mxmpio(8)	Monitor or manage MPIO devices
mxnlmconfig(8)	Enable or disable NLM locking
mxsanlock(8)	Display status of SAN ownership locks
netif(8)	**mx** command to manipulate network interfaces
notifier(8)	**mx** command to manipulate notifiers
psfsck(8)	Check a PSFS filesystem

psfssema(8)	Manage command-line semaphores
psfssuspend(8), psfsresume(8)	Suspend a PSFS filesystem; resume a suspended PSFS filesystem
resizepsfs(8)	Resize a PSFS filesystem
sandiskinfo(8)	Display information about PSFS filesystems
server(8)	**mx** command to manipulate servers
service(8)	**mx** command to manipulate service monitors
sleep(8)	Pause between **mx** commands
snapshot(8)	**mx** command to manage hardware snapshots
vhost(8)	**mx** command to manipulate virtual hosts
vnfs(8)	**mx** command to manage Virtual NFS Services

3

Configure Servers

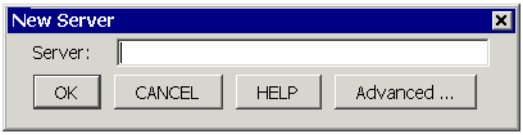
Before adding a server to a cluster, verify the following:

- The server is connected to the SAN if it will be accessing PSFS filesystems.
- The server is configured as a fully networked host supporting the services to be monitored. For example, if you want HP Clustered File System to provide failover protection for your Web service, the appropriate Web server software must be installed and configured on the servers.
- If the */etc/hosts* file has been modified, it should be consistent with the DNS.

Add or Modify a Server

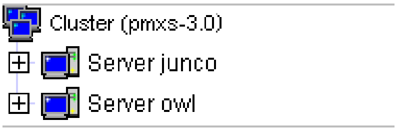
To add a new server to a cluster, select Cluster > Server > New (or click the Add Server icon on the toolbar). Then configure the server on the New Server window.

To modify an existing server, select that server on the Servers window, right-click, and select Properties. Then make your changes on the Server Properties window.



Server: Enter the name or IP address of the server.

The new server appears on the Servers window. In the following example, two servers have been added to a cluster.



Cluster (pmxs-3.0)

⊞ Server junco

⊞ Server owl

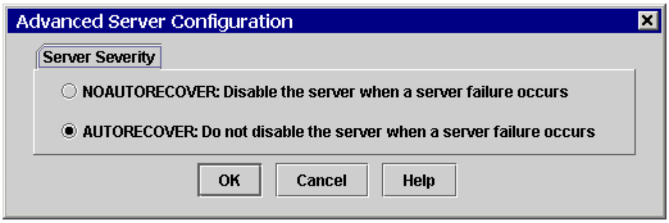
NOTE: For improved performance, the Management Console caches hostname lookups. If your DNS changes, you may need to restart the console so that it will reflect the new hostname.

To add or update a server from the command line, use this command:

```
mx server add|update [--serverSeverity=autorecover|
noautorecover] <server> ...
```

Advanced Settings for Servers

The Advanced button on the New Server or Server Properties window is used to set the failback behavior, also called *Server Severity*, for the server.



Advanced Server Configuration

Server Severity

○ NOAUTORECOVER: Disable the server when a server failure occurs

● AUTORECOVER: Do not disable the server when a server failure occurs

When a server fails completely because of a power failure or other serious event, HP Clustered File System attempts to move any virtual hosts from the network interfaces on the failed server to backup network interfaces on healthy servers in the cluster.

After the server is restored to normal operation (for example, the power is restored and the server is rebooted), it is by default made available to host virtual hosts. If a failure of the server caused failover of virtual hosts, recovery of that server will cause the same set of virtual hosts to automatically “failback” to network interfaces on the original server.

The Server Severity can be configured on each server. There are two settings:

AUTORECOVER. This is the default behavior. Virtual hosts automatically fail back to network interfaces on the original server after it is restored to normal operation.

NOAUTORECOVER. After the server recovers, it is not made available to host virtual hosts. Instead, HP Clustered File System disables the server. (You will need to re-enable the server with the Management Console or **mx** utility).

The **NOAUTORECOVER** setting can be useful when integrating HP Clustered File System with custom applications, where additional actions may be necessary after server recovery and before the server is made available to host services provided by virtual hosts.

To set the server severity from the command line, use this command:

```
mx server add|update --serverSeverity autorecover|
noautorecover <server> ...
```

Other Server Configuration Procedures

Move a Server

If you move a server from one FibreChannel switch port to another, you will need to enable the server's new port on the switch before starting HP Clustered File System on the server.

Delete a Server

Select the server to be deleted from the Servers window on the Management Console, right-click, and select Delete.

To delete servers from the command line, use this command:

```
mx server delete <server> ...
```

Disable a Server

Select the server to be disabled from the Servers window on the Management Console, right-click, and select Disable.

When you disable a server, the network interfaces on that server are also disabled, which can cause virtual hosts to fail over to another network interface.

To disable servers from the command line, use this command:

```
mx server disable <server> ...
```

Enable a Server

Select the server to be enabled from the Servers window on the Management Console, right-click, and select Enable.

To enable servers from the command line, use this command:

```
mx server enable <server> ...
```

Change the IP Address for a Server

A server's IP address can be changed without affecting the other servers in the cluster. The following procedure uses this scenario: three servers, S1, S2, and S3, with a single NIC. You want to change the IP address of server S2.

1. Stop HP Clustered File System on server S2.
2. Change the IP address of server S2. We will now identify the server as S2a.
3. Start HP Clustered File System on server S2a. The server joins the cluster, which now consists of servers S1, S2, S3, and S2a. Server 2 is down and S1, S2a, and S3 are up.
4. Delete server S2 from the cluster. This step will remove references to the server.
5. Update virtual hosts and any other cluster entities that used server S2 to now include S2a.

HP Clustered File System License File

To operate properly, HP Clustered File System requires that a license file be installed on each server in the cluster.

Upgrade the License File

When you receive a new license file, you will need to install it on the cluster servers. You can either copy the file manually to each server, or you can install it from the Management Console.

Re-read the License File

On each server, HP Clustered File System reads the license file upon startup and at 15-minute intervals. If you have upgraded the license file on a server and want HP Clustered File System to read the new license immediately, use the Re-read License feature (select the server on the Management Console and then select Cluster > Server > Re-read License).

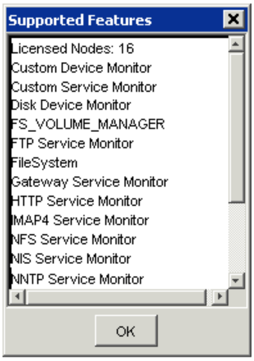
You can also use the following command:

```
mx server read_license <server>
```

Supported HP Clustered File System Features

HP Clustered File System provides device monitors, service monitors, and notifiers. The license agreement for each server determines which features are supported on that server. You can use the Display Features option on the Management Console to determine the supported features for a particular server. Select the server on the Servers window, right-click, and select Display Features.

The Supported Features window lists all of the HP Clustered File System features. Any features not supported by your license are greyed out.



Licensed Nodes: 16
Custom Device Monitor
Custom Service Monitor
Disk Device Monitor
FS_VOLUME_MANAGER
FTP Service Monitor
FileSystem
Gateway Service Monitor
HTTP Service Monitor
IMAP4 Service Monitor
NFS Service Monitor
NIS Service Monitor
NNTP Service Monitor

Migrate Existing Servers to HP Clustered File System

In HP Clustered File System, the names of your servers should be different from the names of the virtual hosts they support. A virtual host can then respond regardless of the state of any one of the servers.

In some cases, the name of an existing server may have been published as a network host before HP Clustered File System was configured. You can either reassign the public name of the server to the virtual host, which is generally the better solution, or keep the published name associated with the server.

- **Migrate the existing server name to a virtual hostname.** In this case, you will need to assign a new name to the server. External sites that then access the original server name will actually be served by the virtual host—basically, a failover-protected version of the server, with no difference in appearance to the clients.
- **Keep the existing name on the server.** If you do not rename the server, clients will need to use the new virtual host name to benefit from failover protection. Clients can still access the server by its name, but those requests are not protected by HP Clustered File System. If the server fails, requests to the server's hostname fail, whereas requests to the new virtual hostname are automatically redirected by HP Clustered File System to a backup server.

Configure Servers for DNS Load Balancing

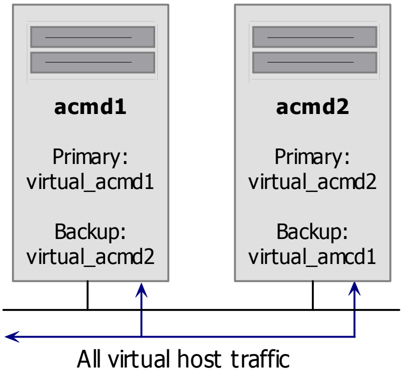
HP Clustered File System can provide failover protection for servers configured to provide domain name service (DNS) load balancing using BIND 4.9 or later. DNS load balancing allows you to set up servers so that requests are sent alternately to each of the servers in a cluster. For example, clients could access acmd1 and acmd2 in turn, enabling the servers to share the load of accepting and responding to client requests.

BIND 4.9 includes a round-robin feature using address records (A records) to alternate requests between IP addresses. This feature provides a straightforward way to gain load balancing capability and simplified load balancing, resulting in improved response time.

HP Clustered File System provides failover protection for this configuration. Without HP Clustered File System, requests are simply alternated between the servers. If a server goes down, requests to that server do not connect.

To configure for round-robin load balancing with HP Clustered File System, you define virtual hosts as addresses in the A records on the DNS. Then use HP Clustered File System to associate primary and backup servers with that virtual host. In essence, you use the virtual host capability of HP Clustered File System to accept the round-robin requests sent by the DNS. HP Clustered File System handles the failover function to the real servers in the cluster.

In the following example, two servers, acmd1 and acmd2, are configured to accept round-robin requests.



The addresses on the name server are *virtual_acmd1* and *virtual_acmd2*. Two virtual hosts have also been created with those names. The first virtual host uses acmd1 as the primary server and acmd2 as the backup. The second virtual host uses acmd2 as the primary and acmd1 as the backup.

The DNS server is configured for round robin using the following A records:

Address	**Time to Live**	**Service**	**Record Type**	**IP Address**
www.acmd.com.	60	IN	A	10.1.1.1
www.acmd.com.	60	IN	A	10.1.1.2

Address: The virtual hostnames that customers use to send requests to your site. (The period following the ".com" in the address is required.)

Time to Live: The number of seconds an address can be cached by intermediate DNS servers for load balancing. This value should be relatively brief so that the addresses of non-functioning servers are not cached by intermediate DNS servers for long periods of time.

Service: Always **IN** for internet.

Record Type: Always **A** for address.

IP address: The IP addresses for the virtual hosts you will use for each server in the cluster. These are the IP addresses that the DNS will use to send alternate requests.

With this setup, the domain name server sends messages in a round-robin fashion to the two virtual hosts indicated by the IP addresses, causing them to share the request load. Each virtual host has a single primary server on which it performs its work; the other server is a backup in case the primary goes down or is taken offline for maintenance.

Depending on your needs, more complicated combinations of DNS round robin and HP Clustered File System can be used, including cluster configurations with multiple failover-protected load-balanced servers.

4

Configure Network Interfaces

When you add a server to the cluster, HP StorageWorks Clustered File System determines whether each network interface on that server meets the following conditions:

- The network interface is up and running.
- Broadcast and multicast are enabled on the network interface.
- Each network interface card (NIC) is on a separate network.

Network interfaces meeting these conditions are automatically configured into the cluster.

HP Clustered File System ignores aliased network interfaces and does not configure them into the cluster.

Overview

Be sure that your network configuration meets the requirements specified in the *HP StorageWorks Clustered File System Setup Guide*.

Administrative Traffic

HP Clustered File System uses a local network for its administrative traffic. It does not require exclusive use of this network; other site traffic can also use it.

By default, HP Clustered File System administrative traffic is allowed on all network interfaces.

However, when you configure the cluster, you can specify the networks that you prefer to use for the administrative traffic. For performance reasons, we recommend that these networks be isolated from the networks used by external clients to access the cluster.

When HP Clustered File System is started, the PanPulse daemon selects the administrative network from the available networks. When a new server joins the cluster, the PanPulse daemon on that server tries to use the established administrative network.

If it cannot use that network, the PanPulse daemon on the new server will look for another network that all of the servers can use. If this attempt fails, the new server will attempt to communicate with the other servers over the main interface (the hostname interface) if possible; however, it may drop out of the cluster.

Network Topology

The cluster should have a symmetrical network configuration; each server should be connected to the same set of networks. This topology helps to ensure that the administrative network can fail over to another network and maintain communications with all servers in the cluster.

If the network topology is asymmetric, where networks are connected to only some of the servers, it may not be possible to fail over the administrative network to another network because the new network might not reach all of the servers in the cluster.

Virtual Hosts

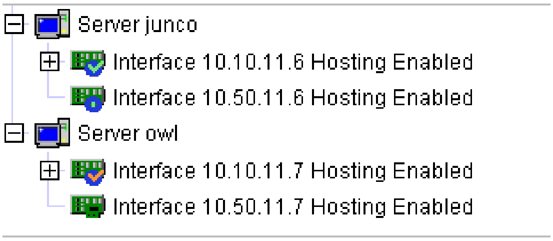
A virtual host is created on a set of network interfaces. These network interfaces must be enabled for virtual hosting. By default, all network interfaces are enabled; however, you can disable a network interface if you do not want it to carry virtual host traffic.

Network Interfaces and the Management Console

When HP Clustered File System is started, the ClusterPulse daemon probes the servers in the cluster to locate network interfaces and then adds the interfaces to its configuration file. The Servers window on the Management Console shows the network interfaces for each server as defined in this file.

(Because there can be stale information in the configuration file, the Servers window may not match your current network configuration exactly.)

Each network interface is labeled “Hosting Enabled” or “Hosting Disabled,” which indicates whether it can be used for virtual hosts.



```
⊟ Server junco
   ⊞ Interface 10.10.11.6 Hosting Enabled
     Interface 10.50.11.6 Hosting Enabled
⊟ Server owl
   ⊞ Interface 10.10.11.7 Hosting Enabled
     Interface 10.50.11.7 Hosting Enabled
```

The Management Console uses the following icons to represent the status of each network interface.

✓ The network interface is currently in use as the administrative network on this server (the green checkmark indicates that the network interface is configured to allow administrative traffic).

✓ The network interface is currently in use as the administrative network on this server (the orange checkmark indicates that the network interface is configured to discourage administrative traffic but no other network interface could be used by all of the servers).

● The network interface allows administrative traffic.

▬ The network interface discourages administrative traffic.

✗ A network interface that allows administrative traffic is down.

✗ A network interface that discourages administrative traffic is down.

⚠ The netmask for this network interface is unknown.

Administrative Network Failover

An administrative network failure occurs when the interface on a particular server is no longer receiving HP Clustered File System administrative traffic. Some possible causes of the failure are a bad cable or network interface card (NIC), or because the administrator disabled the interface with the **ifconfig** command.

When the administrative network fails on a server, the PanPulse daemon on that server attempts to select another network to act as the administrative network. A message regarding the network failure also appears in the *matrix.log* file.

When the PanPulse daemon locates another network that all servers in the cluster can access, all of the servers fail over the administrative network to that network. The daemon looks for another network in this order:

- Networks that allow administrative traffic.
- Networks that discourage administrative traffic.

If HP Clustered File System must use a network that was configured to discourage administrative traffic, it will fail over to a network that allows the traffic as soon as that network becomes available to all servers in the cluster.

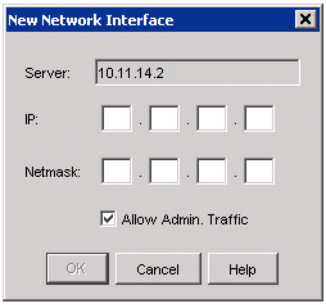
If multiple interface failures occur on a server and there is not another network available for the administrative network, the server may drop out of the cluster. The remaining servers will continue to use the existing administrative network.

Add or Modify a Network Interface

When you add a server to the cluster, its network interfaces are automatically configured into the cluster. Occasionally, you may want to preconfigure a network interface. For example, you can add a network interface that either does not currently exist on a server, or that is on a server that is down. You might also want to modify an existing interface.

- To add a network interface, select the server for that interface on the Servers window, right-click, and select Add Network Interface.

- To modify an existing network interface, select that interface, right-click, and select Properties. The network interface must be down; you cannot modify an "up" network interface.


New Network Interface

Server: 10.11.14.2

IP: □ . □ . □ . □

Netmask: □ . □ . □ . □

☑ Allow Admin. Traffic

Server: The name or IP address of the server that will include the new network interface.

IP: Type the IP address for the network interface.

Net Mask: Type the net mask for the network interface.

Allow Administrative Traffic: Specify whether the network interface can host administrative traffic. The default is to allow the traffic.

NOTE: The definition for a network interface cannot be modified while HP Clustered File System is running. If any part of a network interface definition needs to be changed, stop HP Clustered File System on the affected server, shut down the network, and then make the change. When the change is complete, bring up the network and restart HP Clustered File System.

To add a network interface from the command line, use the following command:

```
mx netif add [--adminTraffic allow|discourage] --server
<server> --netmask <interface_netmask> <interface_ip>
```

To modify a network interface, use this command:

```
mx netif update [--adminTraffic allow|discourage]
--netmask <interface_netmask> <interface_ip>
```

Remove a Network Interface

This option can be useful when performing off-line configuration of a server. To remove a network interface, select that interface on the Servers window, right-click, and then select Delete.

You cannot delete a network interface that is up. If you need to remove a network interface from an online server, first physically remove the corresponding cable from the server. PanPulse will then report that the network interface is down and you can perform the delete operation.

The **mx** command to remove a network interface is as follows:

```
mx netif delete <interface_ip>
```

Allow or Discourage Administrative Traffic

By default, all network interfaces allow administrative traffic. However, you can specify which networks you prefer to use for this traffic.

To allow or discourage administrative traffic on a network interface, select that network interface on the Servers window, right-click, and then select either “Allow Admin. Traffic” or “Discourage Admin. Traffic” as appropriate.

From the command line, use the following command to allow administrative traffic on specific network interfaces:

```
mx netif admin <interface_ip> ...
```

Use the following command to discourage administrative traffic:

```
mx netif noadmin <interface_ip> ...
```

Enable or Disable a Network Interface for Virtual Hosting

By default, all network interfaces are enabled for virtual hosting. To enable or disable a network interface for virtual hosting, select that interface on the Servers window, right-click, and then select either “Enable Hosting” or “Disable Hosting” as appropriate.

From the command line, use the following command to disable virtual hosting:

```
mx netif disable <interface_ip> ...
```

Use the following command to enable virtual hosting:

```
mx netif enable <interface_ip> ...
```

5

Configure the SAN

SAN configuration includes the following:

- Import or deport SAN disks. After a disk is imported, it can be used for PSFS filesystems.
- Change the partitioning on SAN disks.
- Display information about SAN disks.
- Manage multipath I/O.

Overview

SAN Configuration Requirements

Be sure that your SAN configuration meets the requirements specified in the *HP StorageWorks Clustered File System Setup Guide*.

Storage Control Layer Module

The Storage Control Layer (SCL) module manages shared SAN devices. It operates as a daemon (SCLd) running on each server in the cluster and is created by the SANPulse daemon.

The primary function of the SCL module is to ensure that when a server drops out of the cluster unexpectedly, it is physically disabled from further access to the imported SAN storage devices.

This step enables HP Clustered File System to perform any needed filesystem recovery operations to ensure filesystem integrity and then to continue running safely on the remaining set of servers.

As part of managing shared SAN devices, the SCL is also responsible for providing each disk with a globally unique device name that all servers in the cluster can use to access the device.

Device Names

The SCL uses unique device names to control access to shared SAN devices. These names form the pathnames that servers use to access shared data.

When you import a SAN disk, the SCL gives it a global device name that represents the entire disk. The name consists of **psd** followed by an integer. For example, **psd1**, **psd2**, and **psd25** are all valid names.

Individual disk partitions also have a global device name. It consists of the name assigned to the disk followed by **p** and the partition number. For example, partition 6 on disk **psd12** is represented as **psd12p6**.

Global device names are not dependent on the physical location of the disks. If you move a disk that has already been assigned a global device name, it will keep that name after the move.

If you import a disk that already has a device name, the SCL will keep that name if it is not currently in use. If the name has been assigned to another device, the SCL will assign a new name to the disk that you are importing.

Device Database and Membership Partitions

The SCL stores the device name and physical UID for each imported disk in a device database. When you configured HP Clustered File System, you specified a set of SAN partitions to be used for the database and for other functions related to SAN control. These partitions are called *membership partitions* and are automatically imported into the cluster. (For information about creating these partitions, see “Manage Membership Partitions with mxmpconf” on page 183.)

HP Clustered File System creates the device database on one of the membership partitions and then replicates it to the other partitions. The database must be present for a cluster to form. To ensure that the database is always available, the recommended configuration is a three-way mirror.

You can use the **mxmpconf** utility to fix any problems with the membership partitions.

Device Access

Once imported, a shared device can be accessed only with its global device name, such as **psd6p4**. On each server, the SCL creates device node entries in the directory */dev/psd* for every partition on the disk. The names of the entries match the global device names of the partitions. The original */dev* entries for the partitions are disabled, giving HP Clustered File System complete control over all accesses to those shared devices.

Import SAN Disks

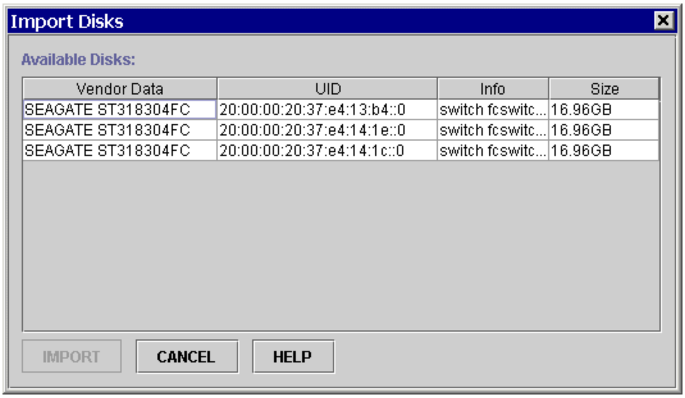
SAN disks to be used for PSFS filesystems must be imported into the cluster. This step gives the cluster control over access to the disks. Before importing disks, you should be aware of the following:

- If servers from multiple clusters can access the SAN via a shared FC fabric, avoid importing the same disk into more than one cluster. Filesystem corruption can occur when different clusters attempt to share the same filesystem.
- If you want to change the partition table on a disk after it has been imported, you will need to deport the disk, make the changes, and then import the disk again.
- Disks containing an active membership partition cannot be imported. You will need to use **mxmpconf** to inactivate the partition (see "Manage Membership Partitions with mxmpconf" on page 183). After the partition is inactivated, you can import the disk into the cluster.
- If you import a large number of disks sequentially, it can take several minutes to complete the imports.

> The imports will be faster if you specify all of the disks in one operation, either by selecting multiple disks on the Import Disks window or by specifying multiple disks on the **mx disk import** command line.

When you import a disk, the SCL gives it a global device name such as **psd25**. It also assigns global device names to all of the partitions on the disk. The individual partitions are identified by the disk name followed by **p** and the partition number, such as **psd25p4**.

To import disks using the Management Console, select Storage > Disk > Import or click the Import Disk icon on the toolbar. The Import Disks window, which appears next, shows all SAN disks that are not currently imported into the cluster. The disk descriptions include the vendor, the disk's UID, and its size. The Info column specifies the FC switch used to access the disk. Select the disks that you want to import from this window.



Vendor Data	UID	Info	Size
SEAGATE ST318304FC	20:00:00:20:37:e4:13:b4::0	switch fcswitc...	16.96GB
SEAGATE ST318304FC	20:00:00:20:37:e4:14:1e::0	switch fcswitc...	16.96GB
SEAGATE ST318304FC	20:00:00:20:37:e4:14:1c::0	switch fcswitc...	16.96GB

To import a disk from the command line, use the following command:

```
mx disk import <uuid> ...
```

To determine the *uuid* for a disk, run the following command, which prints the uuid, the size, and a vendor string for each unimported SAN disk.

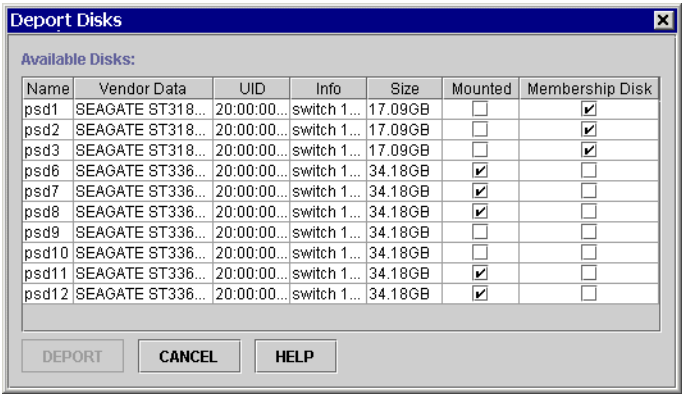
```
mx disk status
```

Deport SAN Disks

Deporting a disk removes it from cluster control. The */dev/psd* device nodes are removed and the original */dev* entries are re-enabled.

You cannot deport a disk that contains a mounted filesystem or a membership partition. Also, disks configured in a dynamic volume cannot be deported. (You will need to destroy the dynamic volume and then deport the disk.)

To deport a disk from the Management Console, select Storage > Disk > Deport or click the Deport Disk icon on the toolbar. The Deport Disks window, which appears next, lists all SAN disks that have been imported into the cluster. Select the disks that you want to deport from this window.



Name	Vendor Data	UID	Info	Size	Mounted	Membership Disk
psd1	SEAGATE ST318...	20:00:00...	switch 1...	17.09GB	☐	☑
psd2	SEAGATE ST318...	20:00:00...	switch 1...	17.09GB	☐	☑
psd3	SEAGATE ST318...	20:00:00...	switch 1...	17.09GB	☐	☑
psd6	SEAGATE ST336...	20:00:00...	switch 1...	34.18GB	☑	☐
psd7	SEAGATE ST336...	20:00:00...	switch 1...	34.18GB	☑	☐
psd8	SEAGATE ST336...	20:00:00...	switch 1...	34.18GB	☑	☐
psd9	SEAGATE ST336...	20:00:00...	switch 1...	34.18GB	☐	☐
psd10	SEAGATE ST336...	20:00:00...	switch 1...	34.18GB	☐	☐
psd11	SEAGATE ST336...	20:00:00...	switch 1...	34.18GB	☑	☐
psd12	SEAGATE ST336...	20:00:00...	switch 1...	34.18GB	☑	☐

To deport a disk from the command line, use the following command:

```
mx disk deport <uuid> ...
```

To determine the `uuid` for the disk, use the **mx disk status --imported** command.

Change the Partitioning on a Disk

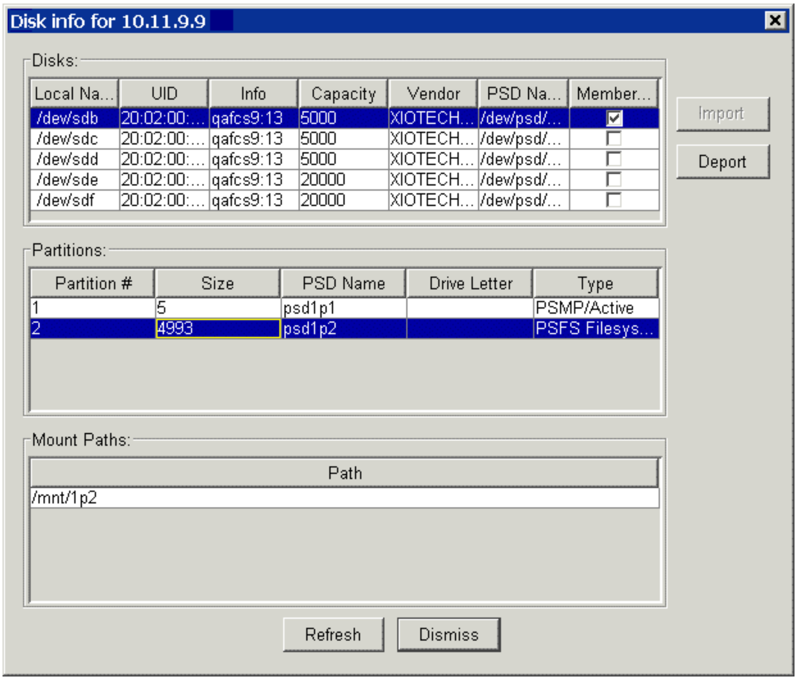
The Linux **fdisk** utility can be used to change the partition layout on a SAN disk. If the disk is currently imported into the cluster, you must first deport the disk.

When you use **fdisk**, the changes made to the partition table are visible only to the server where you made the changes. When you reimport the disk, the other servers in the cluster will see the updated partition table.

Display Local Disk Information

The Disk Info window displays disk information from the viewpoint of the local server. It can be used to match local disk names with the disk names that HP Clustered File System uses (the PSD Name). You can also use this window to import or deport SAN disks.

To see the disk information, select the server on the Servers tab on the Management Console, right-click, and then select Get Local Disk Info.



Disk info for 10.11.9.9

Disks:

Local Na...	UID	Info	Capacity	Vendor	PSD Na...	Member...
/dev/sdb	20:02:00:...	qafcs9:13	5000	XIOTECH...	/dev/psd/...	☑
/dev/sdc	20:02:00:...	qafcs9:13	5000	XIOTECH...	/dev/psd/...	☐
/dev/sdd	20:02:00:...	qafcs9:13	5000	XIOTECH...	/dev/psd/...	☐
/dev/sde	20:02:00:...	qafcs9:13	20000	XIOTECH...	/dev/psd/...	☐
/dev/sdf	20:02:00:...	qafcs9:13	20000	XIOTECH...	/dev/psd/...	☐

Partitions:

Partition #	Size	PSD Name	Drive Letter	Type
1	5	psd1p1		PSMP/Active
2	4993	psd1p2		PSFS Filesys...

Mount Paths:

Path
/mnt/1p2

When you select a disk, the window displays information about the partitions on the disk. Select a partition to display the corresponding Linux mount path for the PSFS filesystem.

To import or deport a disk, select that disk and then click Import or Deport as appropriate.

Display Disk Information with sandiskinfo

The **sandiskinfo** command can display information for both imported and unimported SAN disks and also for dynamic volumes. Under normal operations, the **sandiskinfo** output should be the same on all servers in the cluster.

Disk Information

With no options, **sandiskinfo** displays the UID, vendor, model, and capacity of each imported disk and specifies the FC switch used to access the disk.

```
# sandiskinfo
Disk: /dev/psd/psd5
Uid: 20:00:00:04:cf:13:32:d1::0 SAN info: switch fcswitch port 7
Vendor:      SEAGATE ST336704FC Capacity: 34732M
```

Following are the commonly used options for imported and unimported disks:

```
sandiskinfo [-i|-u|-v] [-al] [-f] [-U]
```

The default is **-i**, which produces the output shown above for imported disks. The **-u** option produces the same output for unimported disks.

The **-U** option displays output in the format used by the Management Console. This option is used internally by HP Clustered File System and does not produce human-readable output.

Show Partition Information

The **-a** option also lists the partitions on each disk. When combined with **-u**, it displays partition information for unimported disks.

```
# sandiskinfo -a
Disk: /dev/psd/psd5
Uid: 20:00:00:04:cf:13:32:d1::0 SAN info: switch fcswitch port 7
Vendor:      SEAGATE ST336704FC Capacity: 34732M
 partition 01: size 3905M type Linux (83)
 partition 02: size  813M type Linux (83) (PSFS Filesystem)
 partition 03: size 7813M type Linux (83) (Membership Partition)
```

Show Local Device Information

The **-l** option displays the local device name for each disk, as well as the default disk information. When combined with **-u**, it displays local device names for unimported disks.

```
# sandiskinfo -al
Disk: /dev/psd/psd5
Uid: 20:00:00:04:cf:13:32:d1::0 SAN info: switch fcswitch port 7
Vendor:      SEAGATE ST336704FC Capacity: 34732M
Local Device Path:  /dev/sda
 partition 01: size 3905M type Linux (83)
 partition 02: size  813M type Linux (83) (PSFS Filesystem)
 partition 03: size 7813M type Linux (83) (Membership Partition)
```

Show Filesystem Information

The **-f** option displays existing PSFS filesystems on imported disks.

```
# sandiskinfo -f
Volume: /dev/psv/psv1            Size:   2439M  (PSFS Filesystem)
    Stripesize=0K
    Local Mount Point=/mnt
Volume: /dev/psd/psd1p6          Size:    490M  (PSFS Filesystem)
    Disk=20:00:00:04:cf:13:38:18::0 partition=06 type=Linux (83)
    Local Mount Point=(not mounted)
```

Show Available Volumes

The **-v** option lists available volumes on imported disks. These volumes are not currently in use for a PSFS filesystem or a membership partition.

```
# sandiskinfo -v
Volume: /dev/psd/psd5p1              Size:  3905M
Disk=20:00:00:04:cf:13:32:d1::0 partition=01 type=Linux (83)
Volume: /dev/psd/psd5p2              Size:  7386M
Disk=20:00:00:04:cf:13:32:d1::0 partition=01 type=Linux (83)
```

Options for Dynamic Volumes

The following **sandiskinfo** options apply only to dynamic volumes.

Show Available Subdevices

The **--subdevices** option lists subdevices that are available for use in constructing a dynamic volume.

```
# sandiskinfo --subdevices
Subdevice: 20:00:00:04:cf:13:38:18::0/2      Size:   1950M   psd1p2
Subdevice: 20:00:00:04:cf:13:38:18::0/7      Size:    490M   psd1p7
Subdevice: 20:00:00:04:cf:13:38:18::0/8      Size:    490M   psd1p8
Subdevice: 20:00:00:04:cf:13:38:18::0/9      Size:    490M   psd1p9
Subdevice: 20:00:00:04:cf:13:38:18::0/10    Size:    490M   psd1p10
Subdevice: 20:00:00:04:cf:13:38:18::0/11    Size:    490M   psd1p11
Subdevice: 20:00:00:04:cf:13:38:18::0/12    Size:    490M   psd1p12
Subdevice: 20:00:00:04:cf:13:38:18::0/13    Size:    490M   psd1p13
Subdevice: 20:00:00:04:cf:13:38:18::0/14    Size:    490M   psd1p14
```

Show Dynamic Volumes

The **--dynvolumes** option lists all dynamic volumes.

```
# sandiskinfo --dynvolumes
Dynamic Volume: psv1             Size:   2439M  Stripe=Unstriped
Dynamic Volume: psv2             Size:    490M  Stripe=32K
Dynamic Volume: psv3             Size:    490M  Stripe=8K
```

Show Properties for Dynamic Volumes

The **--dynvol_properties** [*volname*] option lists detailed properties for the specified dynamic volumes. *volname* is the **psv** name, such as *psv2*. If this option is omitted, the properties for all dynamic volumes are displayed.

```
# sandiskinfo --dynvol_properties
Dynamic Volume: psv1             Size:   2439M  Stripe=Unstriped
   Subdevice: 20:00:00:04:cf:13:38:18::0/5     Size:   490M  psd1p5
   Subdevice: 20:00:00:04:cf:13:38:18::0/2     Size:  1950M  psd1p2
Dynamic Volume: psv2             Size:    490M  Stripe=32K/optimal
   Subdevice: 20:00:00:04:cf:13:38:18::0/7     Size:  490M   psd1p7
Dynamic Volume: psv3             Size:    490M  Stripe=8K/optimal
   Subdevice: 20:00:00:04:cf:13:38:18::0/10    Size:  490M   psd1p10
```

Manage Multipath I/O

HP Clustered File System uses multipath I/O (MPIO) to eliminate single points of failure. A cluster can include multiple FibreChannel switches, multiple FC ports per server, and multiported SAN disks. This configuration provides multiple paths for I/O between cluster servers and SAN disks.

When you start HP Clustered File System, it automatically configures all paths from each cluster server to the storage devices. On each server, it then uses the first path it discovered for I/O with the SAN devices. If that path fails, HP Clustered File System automatically fails over the I/O to another path.

The **mxmpio** command can be used to display status information about MPIO devices or to control the path used for I/O. With the exception of **enableall** and **disableall**, all commands take PSD device names to operate on specific devices. If you do not specify a PSD device, the command operates on all devices.

Enable or Disable Failover for a Server

A cluster server can use multiple FC ports to connect to the FibreChannel fabric. If the FC port or FC switch currently being used for I/O should fail, HP Clustered File System will by default fail over the I/O to another FC port/FC switch path.

You can use the following command to control whether this failover behavior can occur on a particular node. Run the command on the server where you want to change the failover behavior. (HP Clustered File System starts with failover enabled.)

```
# mxmpio enableall|disableall
```

Enable or Disable Failover for a PSD Device

When a failure occurs in the I/O path to a particular PSD device, HP Clustered File System by default fails over to another I/O path. You can use this command to control whether this behavior can occur for specific PSD devices. (HP Clustered File System starts with failover enabled.)

```
# mxmpio enable|disable [<PSD-device ...>]
```

Specify the Path for I/O

If you are troubleshooting problems with an I/O path, you may want to direct the I/O to another I/O path. You might also want to manually balance I/O paths on a per-server basis across the cluster. This is possible because **mxmpio** controls are server specific.

You can use the following command to specify either a particular HBA or a PSD device. HP Clustered File System will then fail over the I/O to the path that includes the specified device. In the command, *PSD-device* is specified by the base name of the device path, such as **psd2p1** (not **/dev/psd/psd2p1**).

```
# mxmpio active <target> <PSD-device>
```

target can be one of the following values:

I A numerical index on the PSD device target array (0..).

M,m A decimal major/minor number identifying the host adapter.

M:m A hexadecimal major/minor number identifying the host adapter.

scsi*N*/*C* A scsi bus ID. *N* is the scsi host adapter ID (0..) and *C* is the scsi host channel number (0..) as defined in */proc/scsi/scsi* and */proc/scsi/*/[0-9]*. If *C* is omitted, zero is assumed.

sd*N***s***P* A device node in */dev*. The value is converted to a major/minor number identifying the corresponding host adapter.

With the exception of *I* (the array index), the value specified is converted to the corresponding host adapter/channel before being used to select the target.

An Example of Changing the I/O Path

In this example, we will change the target for a device. The **mxmpio status -l** command identifies the path currently being used by each device. That path is labeled “active.” The following output shows that device psd2p1 is active on target 1.

```
# /opt/hpcfs/sbin/mxmpio status -l
MPIO Failover is globally enabled
            Failover  Timeout  Targets
    psd1    enabled    30000   0. (41:50) scsi2/0/2/19 (active)
                               1. (08:90) scsi1/0/2/19
    psd1p1  enabled    10000   0. (41:51) scsi2/0/2/19
                               1. (08:91) scsi1/0/2/19 (active)
    psd1p2  enabled    30000   0. (41:52) scsi2/0/2/19 (active)
                               1. (08:92) scsi1/0/2/19
    psd2    enabled    30000   0. (41:10) scsi2/0/1/20 (active)
                               1. (08:50) scsi1/0/1/20
    psd2p1  enabled    10000   0. (41:11) scsi2/0/1/20
                               1. (08:51) scsi1/0/1/20 (active)
    psd2p2  enabled    30000   0. (41:12) scsi2/0/1/20 (active)
                              1. (08:52) scsi1/0/1/20
```

Now use the **mxmpio** command to change the path for psd2p1 to target 0:

```
# /opt/hpcfs/sbin/mxmpio active 0 psd2p1
```

To verify the change, run the **mxmpio status -l** command again. In the following output, device psd2p1 is now active on target 0.

```
# /opt/hpcfs/sbin/mxmpio status -l
MPIO Failover is globally enabled
            Failover  Timeout  Targets
    psd1    enabled    30000   0. (41:50) scsi2/0/2/19 (active)
                               1. (08:90) scsi1/0/2/19
    psd1p1  enabled    10000   0. (41:51) scsi2/0/2/19
                               1. (08:91) scsi1/0/2/19 (active)
    psd1p2  enabled    30000   0. (41:52) scsi2/0/2/19 (active)
                               1. (08:92) scsi1/0/2/19
    psd2    enabled    30000   0. (41:10) scsi2/0/1/20 (active)
                               1. (08:50) scsi1/0/1/20
    psd2p1  enabled    10000   0. (41:11) scsi2/0/1/20 (active)
                               1. (08:51) scsi1/0/1/20
    psd2p2  enabled    30000   0. (41:12) scsi2/0/1/20 (active)
                              1. (08:52) scsi1/0/1/20
```

Display Status Information

The **status** command displays MPIO status information, including the timeout value, whether MPIO is enabled (globally and per-device), and any targets specified with the **active** command. Use the **-l** option to display more information about the targets, as in the above example.

```
$ mxmpio status
```

In contrast with the earlier example, this example shows output for a system that is not MPIO:

```
# /opt/hpcfs/sbin/mxmpio status -l
MPIO Failover is globally enabled
            Failover  Timeout  Targets
    psd1    enabled    30000   0. (41:50) scsi2/0/2/19
    psd1p1  enabled    10000   0. (41:51) scsi2/0/2/19
    psd1p2  enabled    30000   0. (41:52) scsi2/0/2/19
    psd2    enabled    30000   0. (41:10) scsi2/0/1/20
    psd2p1  enabled    10000   0. (41:11) scsi2/0/1/20
    psd2p2  enabled    30000   0. (41:12) scsi2/0/1/20
```

Display MPIO Statistics

mxmpio also includes the following commands, which display various MPIO statistics. See the **mxmpio**(8) man page for details about these commands.

mpiostat Display the number of transient errors for each target, and show the number of failovers and fatal errors for each device.

mpioload Display load information for each target and the PSD device, and show the number of failovers and fatal errors for each device.

iostat Display general I/O statistics for each device.

rawstat Dump general I/O statistics for each device in raw format. The output is intended for use in programs.

Set the Timeout Value

The default timeout period for PSD devices is 30 seconds. If you need to modify this value for a particular PSD device, use the following command. *value* is in milliseconds; however, the smallest unit is 10 milliseconds. A value of zero disables timeouts.

```
# mxmpio timeout value [PSD-device]
```

6

Configure Dynamic Volumes

HP Clustered File System includes a Clustered Volume Manager that you can use to create, extend, recreate, or destroy dynamic volumes, if you have purchased the separate license. Dynamic volumes allow large filesystems to span multiple disks, LUNs, or storage arrays.

Overview

Basic and Dynamic Volumes

Volumes are used to store PSFS filesystems. There are two types of volumes: dynamic and basic.

Dynamic volumes are created by the Clustered Volume Manager. They can include one or more disks, disk partitions, or LUNs that have been imported into the cluster. These entities are called *subdevices*. A single PSFS filesystem can be placed on each dynamic volume. Additional subdevices can be added to a dynamic volume as necessary, and the filesystem on the volume can also be extended.

A basic volume consists of a single disk, disk partition, or LUN that has been imported into the cluster. A PSFS filesystem is then created directly on the disk, partition, or LUN. Basic volumes can be converted to dynamic volumes.

Types of Dynamic Volumes

HP Clustered File System supports two types of dynamic volumes: striped and concatenated. The volume type determines how data is written to the volume.

- **Striping.** When a dynamic volume is created with striping enabled, a specific amount of data (called the stripe size) is written to each subdevice in turn. For example, a dynamic volume could include three subdevices and a stripe size of 64 KB. That amount of data will be written to the first subdevice, then to the second subdevice, and then to the third subdevice. This method fills the subdevices at the same rate and may provide better performance.
- **Concatenation.** When a dynamic volume is created without striping, data is written to the first subdevice until it is full. Then data is written to the second subdevice, and so on.

Dynamic Volume Names

The Clustered Volume Manager uses unique device names to control access to dynamic volumes. These names form the pathnames that servers use to access shared data.

When you create a dynamic volume, the Clustered Volume Manager gives it a global device name. The name consists of **psv** followed by an integer. For example, **psv1**, **psv2**, and **psv25** are all valid names.

HP Clustered File System stores information about dynamic volumes in a volume database located on the membership partitions.

Configuration Limits

The configuration limits for dynamic volumes are as follows:

- A maximum size of 16 TB for a dynamic volume.
- A maximum of 512 dynamic volumes per cluster.
- A maximum of 128 subdevices per dynamic volume.
- Concatenated dynamic volumes can be extended up to 128 times; however, the total number of subdevices cannot exceed 128.

- Striped dynamic volumes can be extended up to 16 times; however, the total number of subdevices cannot exceed 128.

Guidelines for Creating Dynamic Volumes

When creating striped dynamic volumes, follow these guidelines:

- The subdevices used for a striped dynamic volume should be the same size. The Clustered Volume Manager uses the same amount of space on each subdevice in the stripset. When a striped dynamic volume is created, the Clustered Volume Manager determines the size of the smallest specified subdevice and then uses only that amount of space on each subdevice. If the subdevices are different sizes, the additional space on the larger subdevices will be wasted.

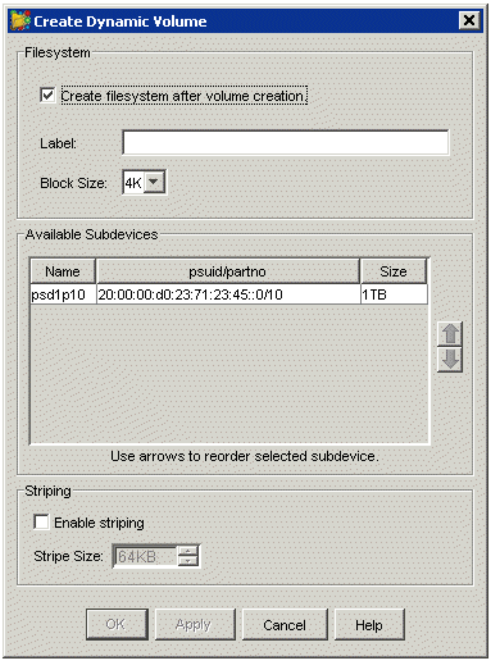
  This limitation also applies when a striped dynamic volume is extended. The newly added subdevices constitute a new stripeset. The size of the smallest subdevice determines the amount of space that will be used on each of the subdevices.

- The default stripe size should be adequate for most applications and hardware configurations. However, be sure to check your application documentation for recommendations regarding the stripe size. Also consider the requirements and performance of the underlying storage hardware.

Create a Dynamic Volume

When you create a dynamic volume, you will need to select the subdevices to be included in the volume. If the volume will be striped, you will also need to select a stripe size. Optionally, HP Clustered File System can also create a filesystem that will be placed on the dynamic volume.

To create a dynamic volume from the Management Console, select Storage > Dynamic Volume > Create Volume.



Filesystem

☑ Create filesystem after volume creation

Label:

Block Size: 4K

Available Subdevices

Name	psuid/partno	Size
psd1p10	20:00:00:d0:23:71:23:45::0/10	1TB

Use arrows to reorder selected subdevice.

Striping

☐ Enable striping

Stripe Size: 64KB

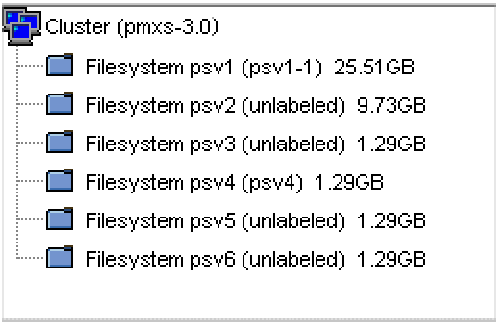
Filesystem: If you want HP Clustered File System to create a filesystem that will be placed on the dynamic volume, enter a label for the filesystem and select the block size. If you do not want a filesystem to be created, remove the checkmark from "Create filesystem after volume creation."

Available Subdevices: The display includes all imported subdevices that are not currently in use by another volume and that do not have a filesystem in place. The subdevices that you select will be used in the order in which they appear on the list. Use the arrow keys to reorder, one-at-a-time, the appropriate subdevices and then highlight those subdevices.

Striping: If you want this volume to use striping, check the "Enable striping" checkbox and then select the size of the stripe. Although the default stripe size of 64KB should be adequate for most applications and hardware configurations, you may need to use a different stripe size for your particular circumstances.

Click either OK or APPLY to create the volume. (APPLY keeps the window open, allowing you to create additional dynamic volumes.)

When the dynamic volume has been created, the Management Console reports the **psv** name assigned to the volume. On the Management Console, filesystems are identified with the **psv** name of the dynamic volume on which they are located, as shown below.



Cluster (pmxs-3.0)
- Filesystem psv1 (psv1-1) 25.51GB
- Filesystem psv2 (unlabeled) 9.73GB
- Filesystem psv3 (unlabeled) 1.29GB
- Filesystem psv4 (psv4) 1.29GB
- Filesystem psv5 (unlabeled) 1.29GB
- Filesystem psv6 (unlabeled) 1.29GB

To create a dynamic volume from the command line, use the following command:

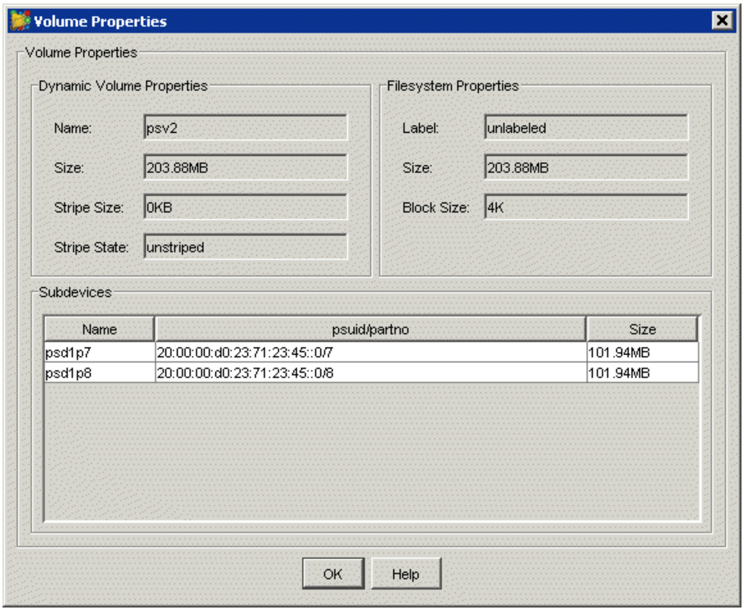
```
mx dynvolume create [--stripesize <4KB-64MB>] <subDeviceName,
subDeviceName,...>
```

The following command lists the available subdevices:

```
mx dynvolume showcreateopt
```

Dynamic Volume Properties

To see the configuration for a dynamic volume, select Storage > Dynamic Volume > Volume Properties and then choose the volume that you want to view. If a filesystem is on the dynamic volume, you can also select that filesystem, right-click, and select Volume Properties. If a filesystem is associated with the volume, the Volume Properties window shows information for both the dynamic volume and the filesystem.



Name: psv2

Size: 203.88MB

Stripe Size: 0KB

Stripe State: unstriped

Label: unlabeled

Size: 203.88MB

Block Size: 4K

Name	psuid/partno	Size
psd1p7	20:00:00:d0:23:71:23:45::0/7	101.94MB
psd1p8	20:00:00:d0:23:71:23:45::0/8	101.94MB

The Stripe State reported in the “Dynamic Volume Properties” section will be one of the following:

- Unstriped. The volume is concatenated and striping is not in effect.
- Optimal. The volume has only one stripeset that includes all subdevices. Each subdevice is written to in turn.
- Suboptimal. The volume has been extended and includes more than one stripeset. The subdevices in the first stripeset will be completely filled before writes to the next stripeset begin. To change the Stripe State to optimal, you will need to recreate the dynamic volume.

To display the properties from the command line, use the following command:

```
mx dynvolume properties <volumeName>
```

Extend a Dynamic Volume

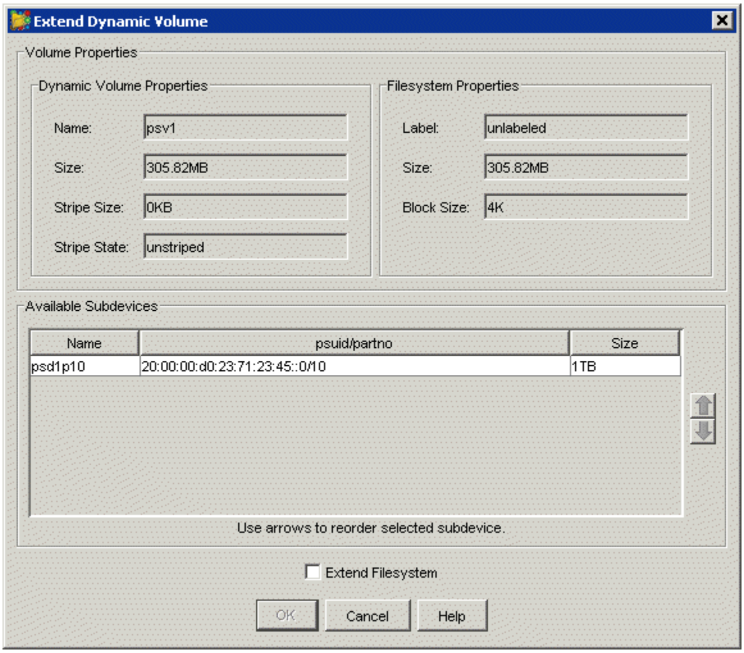
The Extend Volume option allows you to add subdevices to an existing dynamic volume. When you extend the volume on which a filesystem is mounted, you can optionally increase the size of the filesystem to fill the size of the volume.

NOTE: The subdevices used for a striped dynamic volume are called a *stripeset*. When a striped dynamic volume is extended, the new subdevices form another stripeset. The original stripeset is filled first; then the new stripeset is filled. If you want the entire dynamic volume to be in the same stripeset, you will need to recreate the volume.

To extend a dynamic volume on the Management Console, use one of these methods:

- Select Storage > Dynamic Volume > Extend Volume and then choose the volume that you want to extend.
- Select the filesystem located on the volume, right-click, and select Extend Volume.

If a filesystem is on the volume, the Extend Dynamic Volume window shows information for both the dynamic volume and the filesystem.



Name: psv1
Size: 305.82MB
Stripe Size: 0KB
Stripe State: unstriped

Label: unlabeled
Size: 305.82MB
Block Size: 4K

Name	psuid/partno	Size
psd1p10	20:00:00:d0:23:71:23:45::0/10	1TB

Use arrows to reorder selected subdevice.

□ Extend Filesystem

Dynamic Volume Properties: The current properties of this dynamic volume.

Filesystem Properties: The properties for the filesystem located on this dynamic volume.

Available Subdevices: Select the additional subdevices to be added to the dynamic volume. Use the arrow keys to reorder those subdevices if necessary.

Extend Filesystem: To increase the size of the filesystem to match the size of the extended volume, click this checkbox.

When you click OK, the dynamic volume will be extended.

To extend a dynamic volume from the command line, use the following command:

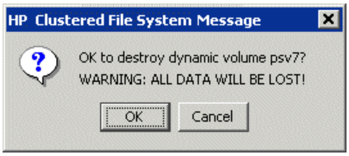
```
mx dynvolume extend <volumeName> <subDeviceName,
subDeviceName,...>
```

Destroy a Dynamic Volume

When a dynamic volume is destroyed, the filesystem on that volume is also destroyed. Before destroying a dynamic volume, be sure that the filesystem is no longer needed or has been copied or backed up to another location. The filesystem must be unmounted when you perform this operation.

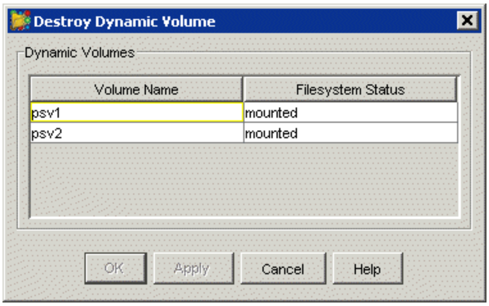
To destroy a dynamic volume from the Management Console, use one of these methods:

- Select the filesystem located on the volume, right-click, and select Destroy Volume. The following warning will appear; when you click OK, the dynamic volume is destroyed.



OK to destroy dynamic volume psv7?
WARNING: ALL DATA WILL BE LOST!

- Select Storage > Dynamic Volume > Destroy Volume. On the Destroy Dynamic Volume window, select the volume that you want to destroy and click either OK or APPLY.



Volume Name	Filesystem Status
psv1	mounted
psv2	mounted

To destroy a dynamic volume from the command line, use the following command:

```
mx dynvolume destroy <volumeName>
```

Recreate a Dynamic Volume

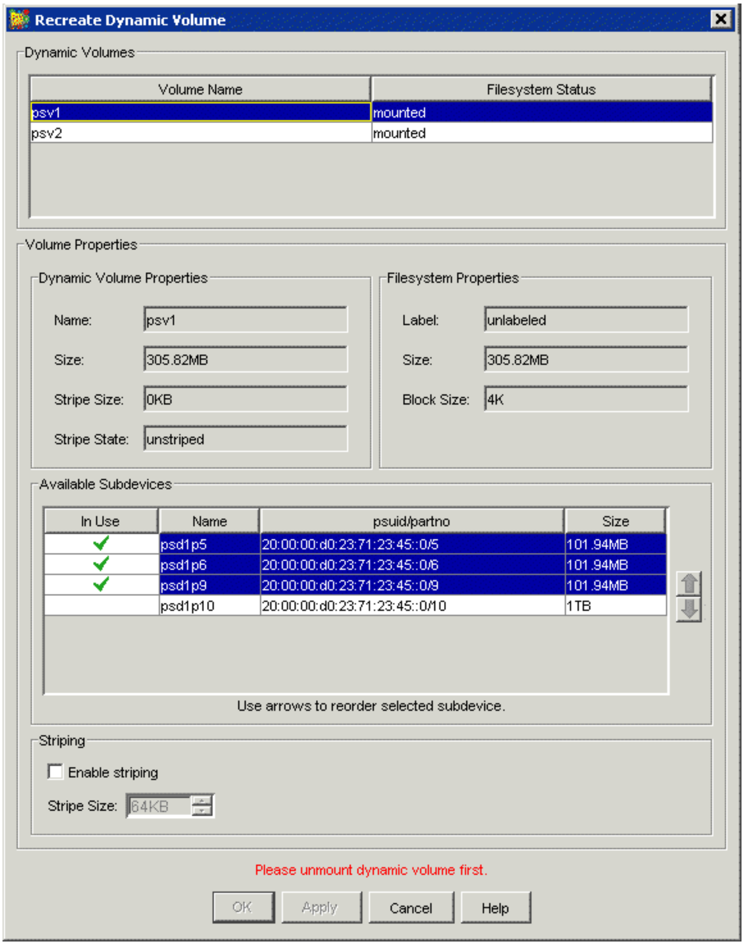
Occasionally you may want to recreate a dynamic volume. For example, you might want to implement striping on a concatenated volume or, if a striped dynamic volume has been extended, you might want to recreate the volume to place all of the subdevices in the same stripe set.

When a dynamic volume is recreated, the Clustered Volume Manager first destroys the volume and then creates it again using the subdevices and options that you select. When the volume is destroyed, any filesystem on that volume is also removed. Before recreating a volume, be sure that the filesystem is no longer needed or has been copied or backed up to another location. The filesystem must be unmounted when you recreate the volume.

To recreate a dynamic volume on the Management Console, use one of these methods:

- Select Storage > Dynamic Volume > Recreate and then choose the volume that you want to recreate.
- If there is a filesystem on the volume, select the filesystem, right-click, and select Recreate Volume.

If a filesystem is mounted on the volume, the Recreate Dynamic Volume window shows information for both the dynamic volume and the filesystem.



Dynamic Volumes

Volume Name	Filesystem Status
psv1	mounted
psv2	mounted

Volume Properties

Dynamic Volume Properties

Name: psv1
Size: 305.82MB
Stripe Size: 0KB
Stripe State: unstriped

Filesystem Properties

Label: unlabeled
Size: 305.82MB
Block Size: 4K

Available Subdevices

In Use	Name	psuid/partno	Size
✓	psd1p5	20:00:00:d0:23:71:23:45::0/5	101.94MB
✓	psd1p6	20:00:00:d0:23:71:23:45::0/6	101.94MB
✓	psd1p9	20:00:00:d0:23:71:23:45::0/9	101.94MB
	psd1p10	20:00:00:d0:23:71:23:45::0/10	1TB

Use arrows to reorder selected subdevice.

Striping

□ Enable striping

Stripe Size: 64KB

Please unmount dynamic volume first.

You can change or reorder the subdevices used for the volume and enable striping if desired.

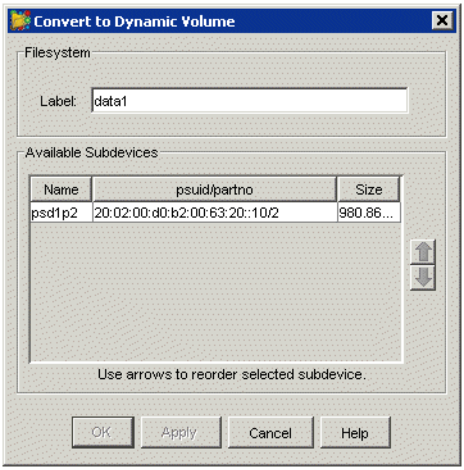
To recreate a volume from the command line, you will first need to use the **dynvolume destroy** command and then run the **dynvolume create** command.

Convert a Basic Volume to a Dynamic Volume

If you have PSFS filesystems that were created directly on an imported disk partition or LUN (a basic volume), you can convert the basic volume to a dynamic volume. The new dynamic volume will contain only the original subdevice; you can use the Extend Volume option to add other subdevices to the dynamic volume.

NOTE: The new dynamic volume is unstriped. It is not possible to add striping to a converted dynamic volume. If you want to use striping, you will need to recreate the volume.

To convert a basic volume, select the associated PSFS filesystem on the Management Console, right-click, and select Convert to Dynamic Volume.



Label: data1

Name	psuid/partno	Size
psd1p2	20:02:00:d0:b2:00:63:20::10/2	980.86...

Use arrows to reorder selected subdevice.

To convert a basic volume to a dynamic volume from the command line, use the following command:

```
mx dynvolume convert <filesystem>
```

7

Configure PSFS Filesystems

HP StorageWorks Clustered File System provides the PSFS filesystem. This direct-access shared filesystem enables multiple servers to concurrently read and write data stored on shared SAN storage devices. A journaling filesystem, PSFS provides live crash recovery.

Overview

Filesystem Features

Concurrent Access by Multiple Servers

After a PSFS filesystem has been created, all cluster servers having access to the device via the SAN can mount the filesystem and subsequently perform concurrent read and/or write operations to that filesystem. A PSFS filesystem must be consistently mounted either read-only or read-write across the cluster.

Standard Filesystem Operations and Semantics

The PSFS filesystem is a standard Linux filesystem and supports standard filesystem operations such as **mkfs**, **mount**, and **umount**. These operations can be performed either with the Management Console or from the command line.

The PSFS filesystem also supports existing applications. It uses standard read/write semantics and does not require any changes to applications.

The PSFS filesystem does not migrate processes from one server to another. If you want processes to be spread across servers, you will need to take the appropriate actions. For example, if you want to spread a kernel build across four servers, you will need to run a cooperative **make**.

Journaling Filesystem

When you initiate certain filesystem operations such as creating, opening, or moving a file or modifying its size, the filesystem writes the metadata, or structural information, for that event to a transaction journal. The filesystem then performs the operation.

If a server using the filesystem should crash during an operation, the journal is replayed and any journaled operations in progress at the time of the crash are completed. This procedure ensures that the structure of the filesystem is consistent. Users on other servers will experience only a slight delay in filesystem operations during the recovery.

Server Registry

HP Clustered File System uses a registry to control server access to PSFS filesystems on the SAN disks imported into the cluster. A server is added to the registry when it mounts its first PSFS filesystem. Under normal operations, a server remains in the registry until it either unmounts its last PSFS filesystem or leaves the cluster.

However, if a server crashes or loses communication with the cluster and that server is in the registry (indicating that it had active PSFS mounts), HP Clustered File System will remove the server’s access to the filesystems. When this situation occurs, an alert will appear on the Management Console and a message will be written to the cluster log. You will need to reboot the server to restore its access.

After the server is rebooted and rejoins the cluster, it is deleted from the registry. When the server again mounts a PSFS filesystem, it will be added back into the registry.

Filesystem Management and Integrity

HP Clustered File System uses the SANPulse daemon to manage PSFS filesystems. SANPulse performs the following tasks.

- Coordinates filesystem mounts, unmounts, and crash recovery operations.
- Checks for cluster partitioning, which can occur when cluster network communications are lost but the affected servers can still access the SAN. If this situation occurs, HP Clustered File System determines which servers should continue to have SAN access and excludes the other servers from the SAN. This step prevents data corruption from occurring.
- Communicates filesystem status to the Management Console. It reports both the global status of a filesystem (for example, whether it is up or going through a recovery operation) and the status of the filesystem on each server (for example, whether the filesystem is mounted or unmounted).

Filesystem Synchronization and Device Locking

The Distributed Lock Manager (DLM) daemon coordinates access to shared resources in the cluster. To do this, it provides the locking mechanisms used to implement the necessary synchronization for various filesystem operations. (Filesystem operations include, but are not limited to, filesystem mounts, reads and writes of files, and the initial creation of a filesystem.)

The operating system kernel uses the DLM to obtain the appropriate lock(s) for all access and modification of files within a PSFS filesystem, ensuring that all PSFS filesystem data are coherent across the cluster. Similarly, the PSFS operations to create, mount, or fsck a filesystem are mutually exclusive, and also use the DLM locking mechanisms.

The DLM provides the fundamental locking mechanisms to prevent mutually exclusive operations from occurring at the same time within the cluster, while allowing operations that can be performed concurrently to proceed in parallel.

Crash Recovery

When a server using a PSFS filesystem either crashes or stops communicating with the cluster, another server in the cluster will replay the filesystem journal to complete any transactions that were in progress at the time of the crash.

Users on the remaining servers will notice a slight delay while the journal is replayed. Typically the recovery procedure takes only a few seconds.

The recovery process restores only the structural metadata information. Any regular file data in the blocks that were in use when the crash occurred may be corrupted or lost.

The server that crashed or lost communication with the cluster will not be able to access the filesystem until it is rebooted.

Filesystem Restrictions

The following restrictions apply to the PSFS filesystem:

- A PSFS filesystem cannot be used as the root or */boot* filesystem.
- A server can mount another non-shared filesystem on a directory of a PSFS filesystem; however, that filesystem will be local to the host. It will not be mounted on other hosts in the cluster.
- PSFS filesystems cannot be mounted using the Linux loop device. These filesystems can be accessed only as SCL-managed “psd” or “psv“ devices.
- The loopback driver cannot be used in conjunction with files located on a PSFS filesystem.
- Swap files are not supported.
- Disk quotas are not supported.
- Inode “last access time” tracking (the **mount -o atime** option) is not supported.
- Certain **fcntl** APIs for user-level applications are not supported. This includes F_GETLEASE, F_SETLEASE, F_NOTIFY, and mandatory locks.

- All options of the **mmap** system call are supported. However, when using PROT_WRITE/MAP_SHARED **mmap**s, cross-server coherency of the mapped regions is not guaranteed. All processes using writable shared **mmap**s to share data must be on the same server.
- The ETXTBSY mechanism is not supported. Processes will not be prevented from writing to running executable files.

Create a Filesystem

A PSFS filesystem can be created on a storage extent located on an imported disk. For the current release, storage extents are basic volumes (**psd** devices) or dynamic volumes (**psv** devices). The maximum filesystem size is 16 TB, which requires a dynamic volume. PSFS filesystems use 4 KB as the block size.

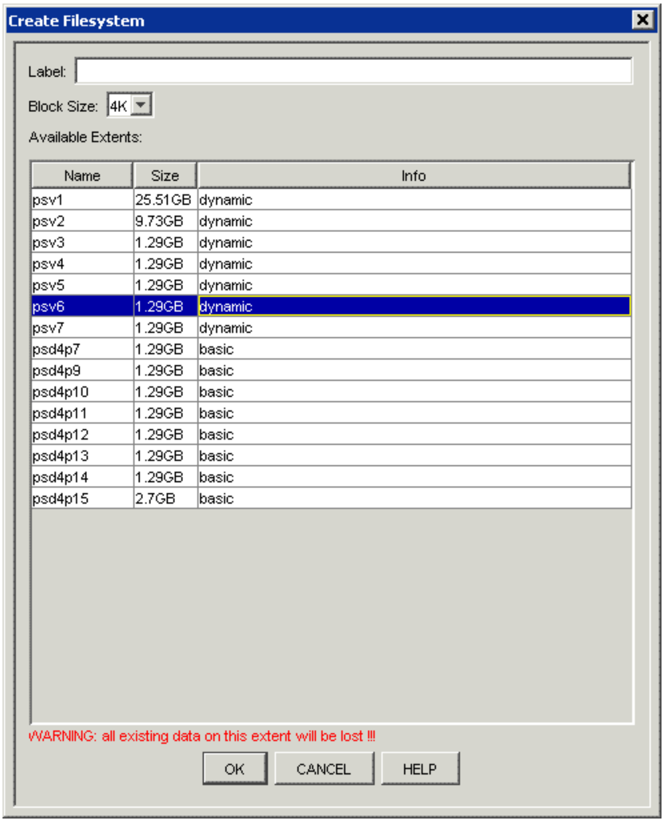
You can create a filesystem from one server in the cluster using either the Management Console or the command line. The filesystem can then be mounted on all servers in the cluster that can access it via the SAN.

Before creating a filesystem, verify the following:

- If the filesystem will be created on a disk partition, ensure that the disk is partitioned appropriately. If you want to change the partition layout, you will need to deport the disk, modify the partition layout, and then reimport the disk.
- The storage extent to be used for the filesystem does not contain needed data. The new filesystem will write over an existing filesystem currently on the storage extent.

Create a Filesystem from the Management Console

To create a filesystem, select Storage > Filesystem > Create Filesystem or click the Add Filesystem icon on the toolbar.



Label:

Block Size: 4K

Available Extents:

Name	Size	Info
psv1	25.51GB	dynamic
psv2	9.73GB	dynamic
psv3	1.29GB	dynamic
psv4	1.29GB	dynamic
psv5	1.29GB	dynamic
psv6	1.29GB	dynamic
psv7	1.29GB	dynamic
psd4p7	1.29GB	basic
psd4p9	1.29GB	basic
psd4p10	1.29GB	basic
psd4p11	1.29GB	basic
psd4p12	1.29GB	basic
psd4p13	1.29GB	basic
psd4p14	1.29GB	basic
psd4p15	2.7GB	basic

WARNING: all existing data on this extent will be lost !!!

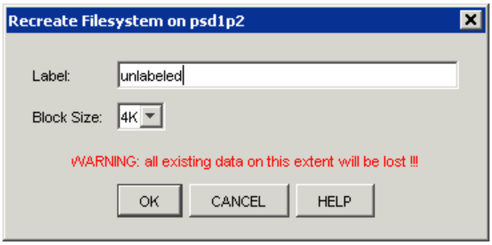
Label: Type a label that identifies the filesystem.

Block Size: The block size is 4K.

Available Extents: The Create Filesystem window lists the storage extents (in this release, basic or dynamic volumes) that are currently unused. Select one of these storage extents for the filesystem.

Recreate a Filesystem

If you want to recreate a filesystem, select the filesystem on the Filesystems window, right-click, and select Recreate Filesystem. All existing data will be lost.



Recreate Filesystem on psd1p2

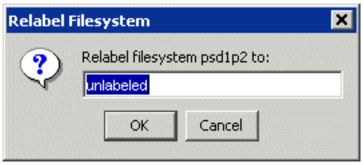
Label: unlabeled

Block Size: 4K

WARNING: all existing data on this extent will be lost !!!

Relabel a Filesystem

To change the identifying label for a filesystem, select the filesystem on the Filesystems window, right-click, and select Relabel Filesystem. You can then enter the new label.



Relabel Filesystem

Relabel filesystem psd1p2 to:

unlabeled

Create a Filesystem from the Command Line

To create a filesystem, use one of the following HP Clustered File System commands.

The mx Command

Use this syntax:

```
mx fs create [--size <KB>] <filesystem> <storageExtent>
```

storageExtent is the basic or dynamic volume where the filesystem will be created. Use the **mx fs showcreateopt** command to locate an available volume. **--size** is optional, and is the number of blocks in the filesystem.

The mkpsfs Command

Use this syntax:

```
mkpsfs -l <label> <device> [<size-in-blocks>]
```

The **-l** option specifies a label for the filesystem.

device is a psd or psv device and is specified as follows:

- For a psd device partition, the device is specified as **/dev/psd/psd***XXX***p***YY*, where XXX is the drive number and YY is the partition number. For example, **/dev/psd/psd6p4** specifies partition 4 on disk psd6.
- For a non-partitioned psd device, the device is specified as **/dev/psd/psd***XXX*, where *XXX* is the drive number. For example, /**dev/psd/psd5**.
- For a psv device, the device is specified as **/dev/psv/psv***XXX*, where *XXX* is the volume number. For example, **/dev/psv/psv1**.

Mount a Filesystem

You can mount a PSFS filesystem on any server that can access the storage extent over the SAN. The directory mountpoint for the filesystem must exist before the filesystem is mounted.

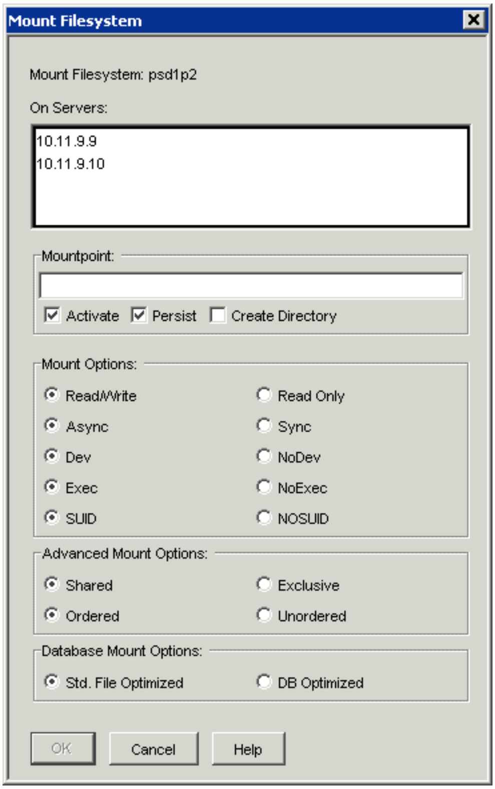
The filesystem must be mounted either read-write or read-only across the cluster. It cannot be mounted read-only on one server and read-write on another server. To change the way the filesystem is mounted, first unmount it on all servers, then remount it.

If you want some servers to have read-write access while other servers have read-only access, mount the filesystem read-write across the cluster. Then, on the servers that should have read-only access, change the permissions on the mountpoint to r-x.

NOTE: The */etc/fstab* file cannot be used to mount PSFS filesystems automatically when the server is rebooted. Instead, mount PSFS filesystems with the “persist” option to cause them to be remounted automatically when the server is rebooted.

Mount a Filesystem from the Management Console

To mount a filesystem, select that filesystem from the Filesystems window, and then either right-click and select Mount, or click the Mount icon on the toolbar.



Mount Filesystem

Mount Filesystem: psd1p2

On Servers:

10.11.9.9
10.11.9.10

Mountpoint:

☑ Activate ☑ Persist □ Create Directory

Mount Options:

● Read/Write ○ Read Only
● Async ○ Sync
● Dev ○ NoDev
● Exec ○ NoExec
● SUID ○ NOSUID

Advanced Mount Options:

● Shared ○ Exclusive
● Ordered ○ Unordered

Database Mount Options:

● Std. File Optimized ○ DB Optimized

Mount Filesystem: This field specifies the location of the filesystem you selected.

On Servers: Select the servers where the filesystem is to be mounted.

Mountpoint: Type the directory mountpoint for the filesystem.

Activate: To mount the filesystem now, click Activate. (This option is enabled by default.)

Persist: This option causes the filesystem to be remounted automatically when the server is rebooted and is enabled by default. If you do not want the filesystem to be remounted automatically, remove the checkmark.

Create Directory: If you want HP Clustered File System to create the directory mountpoint on each server where the filesystem is to be mounted, click Create Directory.

Mount Options: Select the mount options for the filesystem.

- **Read/Write** or **Read Only**. Mount the filesystem read-write or read-only. Read/Write is the default.
- **Async** or **Sync**. Async, the default, allows either asychronous or synchronous I/O to the filesystem. Sync allows only synchronous I/O.
- **Dev** or **NoDev**. Interpret (or do not interpret) character or block special devices on the filesystem. Dev is the default.
- **Exec** or **NoExec**. Permit (or do not permit) the execution of binaries on the mounted filesystem. Exec is the default. Noexec can be used on a system that has filesystems containing binaries for other architectures.
- **SUID** or **NOSUID**. Allow (or do not allow) set-user-id bits and set-group-id bits to take effect. SUID is the default.

Advanced Mount Options: The advanced mount options are as follows:

- **Shared** or **Exclusive**. Either allow all servers having physical access to the filesystem to mount it or allow only one server. Shared is the default.
- **Ordered** or **Unordered**. The Ordered option provides additional security for writes to the filesystem. If a metadata operation will allocate user blocks, the user blocks are written to the filesystem before the metadata is written. With the Unordered option, the writes can occur in either order. The default is Ordered.

Database Mount Option: The database mount option is as follows:

- **Std. File Optimized.** Optimizes file access for standard files.

Mount a Filesystem from the Command Line

To mount a PSFS filesystem from the command line, use one of the following commands.

The HP Clustered File System mx Command

Use this syntax:

```
mx fs mount [--persist] [--activate] [--options
<option>,<option>,...] --path <path> <filesystem> <server>
...
```

The **--persist** argument causes the filesystem to be mounted automatically whenever the server is rebooted. The **--activate** argument mounts the filesystem now. The **--path** argument specifies the directory mountpoint. See the *HP Clustered File System Command Guide Reference* for a description of the options.

The Linux mount Command

Use the following syntax to mount a filesystem. The directory mountpoint must already exist.

```
mount -t psfs -o options device mountpoint
```

The *device* can be either a **psd** or **psv** device. Use this syntax to specify the device:

- For a psd device partition, the device is specified as **/dev/psd/psd***XXX***p***YY*, where XXX is the drive number and YY is the partition number. For example, **/dev/psd/psd6p4** specifies partition 4 on disk psd6.
- For a non-partitioned psd device, the device is specified as **/dev/psd/psd***XXX*, where *XXX* is the drive number. For example, **/dev/psd/psd5**.

- For a psv device, the device is specified as **/dev/psv/psv***XXX*, where *XXX* is the volume number. For example, **/dev/psv/psv1**.

For example, the following command mounts the filesystem on the partition **/dev/psd/psd12p6** in read-write mode at the mountpoint */data1*.

```
mount -t psfs -o rw /dev/psd/psd12p6 /data1
```

The following **mount** options are not supported:

- The **-o atime** option. Inode “last access time” tracking is not supported.
- Use of the **remount** option to change the way a filesystem is mounted. For example, you cannot use **mount -o remount,rw** to remount a filesystem as read-write if it was originally mounted read-only.

The mount operation ignores any options that are not supported by the PSFS filesystem. See **mount_psfs**(8) for more information.

Persistent Mounts

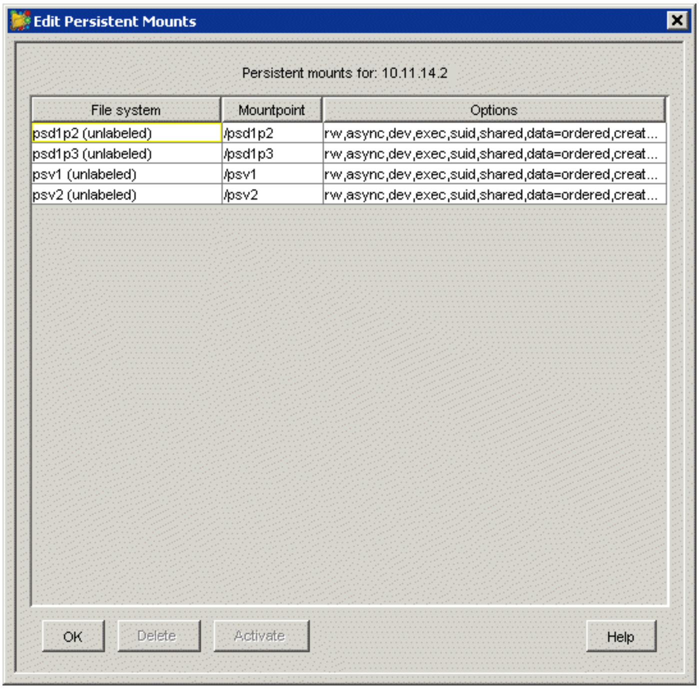
When you mount a filesystem on a server, you can specify that it should be remounted automatically whenever the server is rebooted. When you configure a filesystem mount in this manner, it is a “persistent” mount.

The Edit Persistent Mounts window allows you to remove the “persistent” status from a filesystem and/or to mount the filesystem.

NOTE: If a filesystem was mounted without the “persist” option and you now want it to be persistent, select the mounted filesystem on the Filesystems window, right-click, and select Persist.

Persistent Mounts on a Server

To see all persistent mounts on a particular server, select that server on the Servers window, right-click, and select Edit persistent mount.



Persistent mounts for: 10.11.14.2

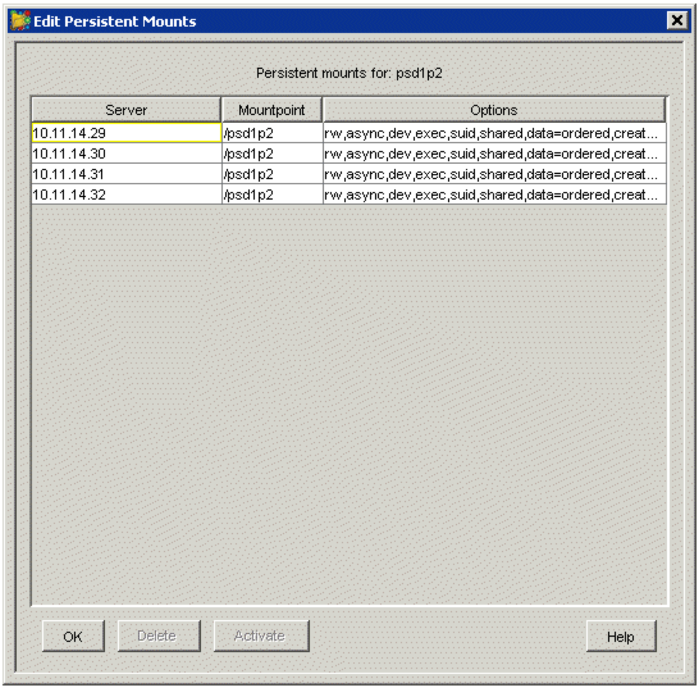
File system	Mountpoint	Options
psd1p2 (unlabeled)	/psd1p2	rw,async,dev,exec,suid,shared,data=ordered,creat...
psd1p3 (unlabeled)	/psd1p3	rw,async,dev,exec,suid,shared,data=ordered,creat...
psv1 (unlabeled)	/psv1	rw,async,dev,exec,suid,shared,data=ordered,creat...
psv2 (unlabeled)	/psv2	rw,async,dev,exec,suid,shared,data=ordered,creat...

The Edit Persistent Mounts window lists all filesystems having a persistent mount on this server.

- To remove the “persistent” mount status for one or more filesystems, select those filesystems and then click Delete.
- To mount a filesystem with the options specified for the persistent mount, select that filesystem and then click Activate.

Persistent Mounts for a Filesystem

To see all persistent mounts for a particular filesystem, select that filesystem on the Filesystems window, right-click, and select Edit persistent mounts.



Persistent mounts for: psd1p2

Server	Mountpoint	Options
10.11.14.29	/psd1p2	rw,async,dev,exec,suid,shared,data=ordered,creat...
10.11.14.30	/psd1p2	rw,async,dev,exec,suid,shared,data=ordered,creat...
10.11.14.31	/psd1p2	rw,async,dev,exec,suid,shared,data=ordered,creat...
10.11.14.32	/psd1p2	rw,async,dev,exec,suid,shared,data=ordered,creat...

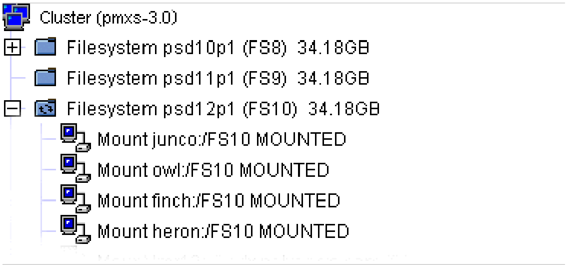
The Edit Persistent Mounts window lists all servers that are configured to have a persistent mount for the filesystem.

- To remove the “persistent” mount status on a particular server, select that server and then click Delete.
- To mount the filesystem with the options specified for the persistent mount, select the appropriate servers and click Activate.

Monitor the Status of Filesystems

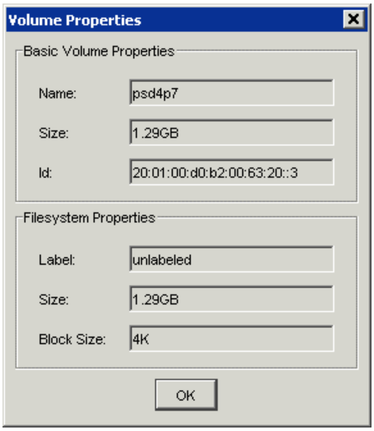
The Filesystems tab on the Management Console lists information about all filesystems in the cluster. If a filesystem is mounted, the entry for that filesystem lists the servers where it is mounted and the mountpoints. The window also shows mount and unmount operations in progress.

The following example shows that filesystem *FS10* is mounted on four servers. The mounts are persistent, as indicated by the icon.



Cluster (pmxs-3.0)
Filesystem psd10p1 (FS8) 34.18GB
Filesystem psd11p1 (FS9) 34.18GB
Filesystem psd12p1 (FS10) 34.18GB
Mount junco:/FS10 MOUNTED
Mount owl:/FS10 MOUNTED
Mount finch:/FS10 MOUNTED
Mount heron:/FS10 MOUNTED

To see information about a specific filesystem, select that filesystem, right-click, and select Volume Properties. The following example shows the properties for a filesystem located on a basic volume.



Volume Properties

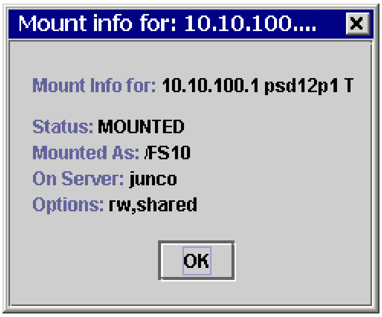
Basic Volume Properties

Name:	psd4p7
Size:	1.29GB
Id:	20:01:00:d0:b2:00:63:20::3

Filesystem Properties

Label:	unlabeled
Size:	1.29GB
Block Size:	4K

To see information about a filesystem mount on a particular server, select the filesystem entry for that server, right-click, and select Properties.



Mount Info for: 10.10.100.1 psd12p1 T

Status: MOUNTED
Mounted As: /FS10
On Server: junco
Options: rw,shared

You can also use Linux utilities such as **df** or the HP Clustered File System **mx** command to display status information.

```
mx fs status [--mounted|--unmounted] [--persistent]
[--verbose] [--standard|--snapshot]
```

The **--verbose** option displays the FS type (always PSFS), the size of the filesystem in KB, and the UUID of the parent disk. The **--standard** option shows only standard filesystems; the **--snapshot** option shows only snapshots.

View Filesystem Errors

The Management Console can display information about the last error that occurred on a filesystem. To see the last error, select the filesystem, right-click, and then select View Last Error.

Unmount a Filesystem

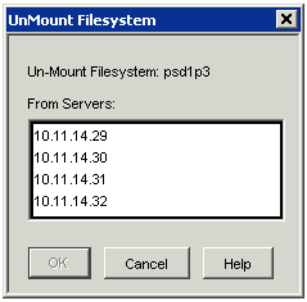
You can unmount a PSFS filesystem from either the Management Console or the command line.

Unmount from the Management Console

The Management Console provides two ways to unmount a filesystem:

- To unmount a filesystem on one server, select the MOUNTED entry for that server on the Filesystems window, right-click, and select Unmount. The filesystem will then be unmounted.
- To unmount a filesystem on more than one server, select the filesystem on the Filesystems window, right-click, and select Unmount.

When you use the second method and select the filesystem, the UnMount Filesystem window is displayed.



UnMount Filesystem

Un-Mount Filesystem: psd1p3

From Servers:

10.11.14.29
10.11.14.30
10.11.14.31
10.11.14.32

UnMount Filesystem: This field lists the filesystem that you selected.

From Servers: Select the servers where the filesystem is to be unmounted.

Unmount from the Command Line

To unmount a filesystem from the command line, use one of the following commands. PSFS filesystems cannot be forcibly unmounted.

The HP Clustered File System **mx** command. The **--persistent** argument removes the persistent status from the filesystem mount; the **--active** argument unmounts the filesystem now.

```
mx fs unmount [--persistent] [--active] <filesys> <server> ...
```

The Linux **umount** command. Be sure to specify the mountpoint, such as */mnt/data1*, not the partition.

```
umount <mountpoint>
```

Other Filesystem Operations

Suspend a Filesystem for Backups

The **psfssuspend** utility suspends a PSFS filesystem in a stable, coherent, and unchanging state. While the filesystem is in this state, you can copy it for backup and/or archival purposes. When copying directly from a suspended device, be sure to use the raw device (*/dev/rpsd/...*) to ensure that all blocks copied are up-to-date.

The filesystem is essentially unusable while it is suspended; however, applications that can tolerate extended waits for I/O do not need to be terminated.

The **psfsresume** utility restores a suspended filesystem. The **psfssuspend** and **psfsresume** utilities affect the specified filesystem on all servers where it is mounted; however, the utilities should be executed on only one server in the cluster.

To suspend a filesystem, issue the following command on one server that has mounted the filesystem. You must be user *root*.

```
# /opt/hpcfs/tools/psfssuspend directory_mountpoint
```

The **psfssuspend** command prevents modifications to the filesystem and forces any changed blocks associated with the filesystem to disk. The command performs these actions on all servers that have mounted the filesystem and then returns successfully.

Any process attempting to modify a suspended filesystem will block until the filesystem is resumed. These blocked processes may hold resources, thereby causing other processes to block waiting on these resources.

When you have completed the backups, use the **psfsresume** utility to resume the suspended filesystem. Issue the command from the server where you executed **psfssuspend**. You must be user *root*.

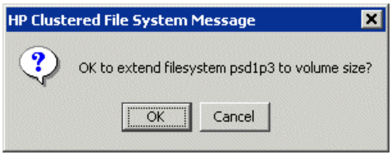
```
# /opt/hpcfs/tools/psfsresume directory_mountpoint
```

NOTE: If an attempt to mount the copied filesystem fails with an "FSID conflict" error, run the following command as user *root*. In the command, <*device*> is the psd or psv device, such as **/dev/psd/psd1p7** or **/dev/psv/psv1**, that contains the copied filesystem, and <*label*> is the name that should be used to identify the filesystem.

```
# /opt/hpcfs/tools/psfslabel <device> "<label>"
```

Extend a Mounted Filesystem

If more space is available on the raw disk or partition on which a filesystem is located, you can extend the filesystem to fill that space. Select the filesystem on the Management Console, right-click, and select Extend Filesystem. You will then see a message such as the following:



HP Clustered File System Message

OK to extend filesystem psd1p3 to volume size?

When you click OK, HP Clustered File System will extend the filesystem if possible. If the filesystem is already using all of the available space, a message will inform you that the filesystem could not be extended.

If the filesystem is on a dynamic volume, you can use the Extend Volume option to increase the size of both the dynamic volume and the filesystem.

Resize a Filesystem Manually

The **resizepsfs** program can be used to increase the size of a PSFS filesystem. **resizepsfs** can grow a filesystem online if the filesystem is mounted on the server where the utility is invoked. Otherwise, the filesystem must be unmounted on all servers before it is resized.

The utility can be used with either **psd** or **psv** devices. Use this syntax to specify the device:

- For a psd device partition, the device is specified as **/dev/psd/psd***XXX***p***YY*, where XXX is the drive number and YY is the partition number. For example, **/dev/psd/psd6p4** specifies partition 4 on disk psd6.
- For a non-partitioned psd device, the device is specified as **/dev/psd/psd***XXX*, where *XXX* is the drive number. For example, **/dev/psd/psd5**.
- For a psv device, the device is specified as **/dev/psv/psv***XXX*, where *XXX* is the volume number. For example, **/dev/psv/psv1**.

This program does not change the size of the partition containing the filesystem. Instead, you will need to use a utility specific to your RAID subsystem to modify the size of the partition. You will need to deport the disk containing the filesystem before you modify the partitions.

CAUTION: Be sure to back up your data before using this program.

You can use the **-s** option to specify the new size for the filesystem. If you do not specify the size, the filesystem will grow to the size of the partition. The **-s** option can be used as follows:

- Specify the size in kilobytes, megabytes, or gigabytes:

  ```
  -S size[K|M|G][T]
  ```

- Specify the amount (in kilobytes, megabytes, gigabytes, or terabytes) by which the filesystem should be increased:

  ```
  -s [+|-]size[K|M|G][T]
  ```

The following example increases the size of the filesystem by 1 GB.

```
resizepsfs -s +1G /dev/psd/psd6p4
```

NOTE: If you do not specify any options, **resizepsfs** will try to resize the filesystem to the full size of the partition.

For more information, see the **resizepsfs**(8) man page.

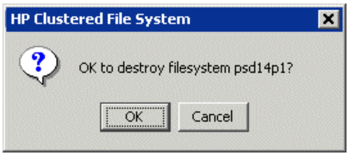
Check a Filesystem

If a filesystem is not unmounted cleanly, the journal will be replayed the next time the filesystem is mounted to restore consistency. You should seldom need to check the filesystem. However, if a filesystem was corrupted by a hardware or software failure, you can repair it with the **psfsck** utility. See the **psfsck**(8) man page for details.

CAUTION: We strongly recommend that you make a backup copy of the entire partition before you attempt to run **psfsck** with the **--rebuild-tree** option. The **--rebuild-tree** option should be used only if **psfsck**, when run in check mode, reports errors that can be fixed only by **--rebuild-tree**. If you run **psfsck** in read-only check mode, we strongly recommend that you run **psfsck** again without the **--no-modify** option.

Destroy a Filesystem

If necessary, you can remove a filesystem from the volume on which it is located. Select the filesystem on the Management Console, right-click, and select Destroy Filesystem.



OK to destroy filesystem psd14p1?

NOTE: Any snapshots of the filesystem must be destroyed before the filesystem is destroyed.

Context Dependent Symbolic Links

A Context Dependent Symbolic Link (CDSL) contains a keyword that identifies a particular location. When the operating system resolves the symbolic link, it translates the keyword into the appropriate pathname. For example, a CDSL could contain the keyword HOSTNAME, which resolves to the hostname of the server where you are logged in.

A CDSL is created with the **ln -s** command. The symbolic link must be located on a PSFS filesystem; however, the target can be on either a PSFS or non-PSFS filesystem.

```
$ ln -s target_with_keyword link_name
```

HP Clustered File System supports the following keywords, which must be enclosed in curly braces.

HOSTNAME	Output from **uname -n** (the node’s network hostname).
MACH	Output from **uname -m** (the machine type).
OS	Output from **uname -s** (the operating system name).
SYS	Concatenated output from **uname -m** and **uname -s**, separated by an underscore (for example, i386_Linux).

UID The effective UID of the process accessing the link. The UID refers to the numerical UID, not the user name.

GID The effective GID of the process accessing the link. The GID refers to the numerical GID, not the group name.

NOTE: CDSLs will not work if they are accessed through NFS because NFS resolves the link on the client.

Examples

Locate a Target by Its Hostname

This example uses three servers: serv1, serv2, and serv3. Each server must have specific configuration files in the */oracle/etc* directory. You can use a CDSL to simplify accessing these server-specific files.

1. Create a subdirectory for each server in */oracle*, which is a PSFS filesystem. Then create an */etc* subdirectory in each server directory. You now have the following directories in the */oracle* PSFS filesystem:

   ```
   /oracle/serv1/etc
   /oracle/serv2/etc
   /oracle/serv3/etc
   ```

2. Populate the directories with the data from the */oracle/etc* directory:

   ```
   cp -r /oracle/etc/* /oracle/serv1/etc/
   cp -r /oracle/etc/* /oracle/serv2/etc/
   cp -r /oracle/etc/* /oracle/serv3/etc/
   ```

3. Edit the files in the server-specific */etc* directories as appropriate.

4. Create the CDSL:

   ```
   ln -s {HOSTNAME}/etc /oracle/etc
   ```

The **ls -l** output looks like this on each server:

```
ls -l /oracle/etc
lrwxrwxrwx 1 root root 14 Jun 16 16:40 /oracle/etc -> {HOSTNAME}/etc
```

If you are logged in on serv1 and access the */oracle/etc* symbolic link, the HOSTNAME keyword resolves to */oracle/serv1/etc.* Similarly, if you are on serv2, the */oracle/etc* symbolic link points to */oracle/serv2/etc*, and on serv3 it points to */oracle/serv3/etc.*

Locate a Target by Its Machine Type

This example uses two servers, serv1 and serv2, that have different machine types. On serv1, the **uname -m** command returns **i386**; on serv2, it returns **alpha**. We need separate */oracle/bin* and */oracle/sbin* directories for each machine type. You can use CDSLs to simplify accessing these machine-specific directories.

1. Create a subdirectory in */oracle* for each machine type and then create a *bin* and *sbin* directory in the new machine-type directories. You now have the following directories in the */oracle* PSFS filesystem:

   ```
   /oracle/i386/bin
   /oracle/i386/sbin
   /oracle/alpha/bin
   /oracle/alpha/sbin
   ```

2. Copy the appropriate binaries to the new *bin* and *sbin* directories.

3. Create CDSLs for */oracle/bin* and */oracle/sbin*:

   ```
   ln -s {MACH}/bin /oracle/bin
   ln -s {MACH}/sbin /oracle/sbin
   ```

The **ls -l** output looks like this on each server:

```
ls -l /oracle/*bin
lrwxrwxrwx  1  root root 10 Jun 16 16:40 /oracle/bin -> {MACH}/bin
lrwxrwxrwx  1  root root 11 Jun 16 16:40 /oracle/sbin -> {MACH}/sbin
```

When you log in on serv1, the i386 machine, the */oracle/bin* and */oracle/sbin* symbolic links resolve to */oracle/i386/bin* and */oracle/i386/sbin*. On serv2, the alpha machine, the links resolve to */oracle/alpha/bin* and */oracle/alpha/sbin.*

Locate a Target That Is Not on a PSFS Filesystem

In the earlier examples, both the symbolic links and the targets were on PSFS filesystems. However, only the symbolic link must be on a PSFS filesystem. The target can be located on a non-PSFS filesystem.

This example uses a CDSL to link the */oracle/etc* directory on the PSFS filesystem to a non-PSFS directory on serv1 or serv2.

1. On each server, create a target directory in */etc* that has the same name as the server's hostname:

   ```
   mkdir /etc/serv1.xvz.com
   mkdir /etc/serv2.xvz.com
   ```

2. Populate the new directories with the appropriate files.
3. Create the CDSL:

   ```
   ln -s /etc/{HOSTNAME} /oracle/etc
   ```

When you are logged in on serv1, the */oracle/etc* symbolic link will point to */etc/serv1.xvz.com*. On serv2, it will point to */etc/serv2.xvz.com.*

Cluster-Wide File Locking

HP Clustered File System supports cluster-wide locks on files located on PSFS filesystems. The locks are implemented with the standard Linux **flock**() system call, which is also known as the BSD flock interface.

The **psfssema** command-line semaphore utility provides a simple synchronization mechanism for managing cluster-wide file locks. The utility can be used in shell scripts on different nodes of a cluster and takes advantage of the PSFS filesystem and its internode communication abilities. For example, you might want to use cluster-wide file locking in a Start or Stop script for a service or device monitor.

Create a Semaphore

To create a PSFS command-line semaphore, use this command:

```
$ psfssema -i <filename>
```

This command creates the PSFS command-line semaphore file <*filename*> and the supplemental file <*filename.pid*> if either file does not already exist. It then initializes the files to create the PSFS command-line semaphore. <*filename*> must be a path on a PSFS filesystem.

Lock a Semaphore

To lock a PSFS command-line semaphore, use this command:

```
$ psfssema -g <filename>
```

This command locks the PSFS command-line semaphore associated with <*filename*>, which is the name of a semaphore file created by **psfssema -i**. If the semaphore is already locked by another node in the cluster, the calling process blocks until the semaphore becomes available.

An error is returned if <*filename*> does not exist or has not been initialized by **psfssema -i**, or if <*filename.pid*> does not exist.

Unlock a Semaphore

To unlock a PSFS command-line semaphore, use this command:

```
$ psfssema -r <filename>
```

The command unlocks the PSFS command-line semaphore associated with <*filename*>, which is a semaphore file created by **psfssema -i**. If other nodes are blocked on the semaphore when **psfssema-r** is called, one of the blocked **psfssema -g** processes will return successfully.

Delete a Semaphore

To delete a PSFS command-line semaphore, use this command:

```
$ psfssema -d <filename>
```

This command deletes the PSFS command-line semaphore associated with <*filename*>, which is the name of a semaphore file created by **psfssema -i.** It also deletes <*filename*> and <*filename.pid*>.

8

Manage Hardware Snapshots

HP Clustered File System provides support for taking hardware snapshots of PSFS filesystems. The subdevices on which the filesystems are located must reside on one or more storage arrays that are supported for snapshots.

Snapshot support can be configured either on the Management Console "Configure Cluster" window or via the **mxconfig** utility. (See the *HP StorageWorks Clustered File System Setup Guide* for more information.) This procedure creates a snapshot configuration file on each server.

Create a Snapshot

A snapshot is a copy-on-write backup of a LUN, volume, or filesystem at an instance in time. Snapshots are commonly used in situations where long-running, serial operations need access to a locked filesystem but locking the filesystem is not feasible.

Snapshots can be both readable and writable and there can often be multiple snapshots existing of a filesystem at a time.

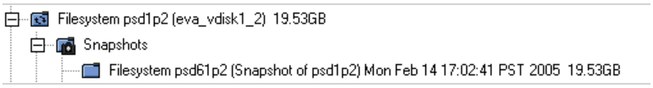
To create a snapshot, select the filesystem on the Management Console, right-click, and select Create Snapshot. (The filesystem must be mounted.) You will then see a dialog asking for information specific to your storage array.

The dialog describes the information that you will need to enter. When you complete the information and click OK, HP Clustered File System takes these steps:

- Quiesces the filesystem to ensure that the snapshot can be mounted cleanly.
- Performs the snapshot operation using the snapshot capability provided by the array. HP Clustered File System selects the next available LUNs for the snapshot.
- Resumes normal filesystem activity.
- Imports the LUNs used for the snapshot into the cluster. The import can take several moments.

When the snapshot is complete, you will be asked whether you want to mount the snapshot. If you choose to do this, the Mount Filesystem window will be displayed. (For information about the mount options, see “Mount a Filesystem” on page 84.)

Mounted snapshots appear on the Management Console beneath the entry for the filesystem. Each snapshot is assigned an HP Clustered File System **psd** or **psv** device name.



Filesystem psd1p2 (eva_vdisk1_2) 19.53GB
Snapshots
Filesystem psd61p2 (Snapshot of psd1p2) Mon Feb 14 17:02:41 PST 2005 19.53GB

To create a snapshot from the command line, first run the following command to determine the options available for the array type on which the specified volume is located:

```
mx snapshot showcreateopt <volume>
```

Then run the following command to create the snapshot:

```
mx snapshot create [--terse] [<options>] <volume>
```

The **--terse** option causes only the name of the snapshot volume to be printed on success.

Errors During Snapshot Operations

If the snapshot operation could not be completed, errors will be reported to the cluster log. You will need to review these errors and may need to delete any partial snapshots by using the Management Appliance.

Delete a Snapshot

Storage arrays typically limit the number of snapshots that can be taken of a specific filesystem. Before taking an additional snapshot, you will need to delete an existing snapshot. Also, if you want to destroy a filesystem, you will first need to delete all snapshots of that filesystem.

To delete a snapshot, select the snapshot on the Management Console, right-click, and select Delete.

To delete a snapshot from the command line, type the following:

```
mx snapshot destroy <volume>
```

Mount or Unmount Snapshots

Snapshots can be mounted or unmounted in the same manner as filesystems. Select the snapshot on the Management Console, right-click, and select Mount or Unmount as appropriate. If you are mounting a snapshot, the Mount Filesystem window appears, allowing you to select mount options.

To mount a snapshot from the command line, type the following:

```
mx fs mount <snapshot>
```

To unmount a snapshot from the command line, type the following:

```
mx fs unmount <snapshot>
```

9

Configure Virtual Hosts

HP Clustered File System uses virtual hosts to provide failover protection for servers and network applications.

Overview

A virtual host is a hostname/IP address configured on a set of network interfaces. Each interface must be located on a different server. The first network interface configured is the primary interface for the virtual host. The server providing this interface is the primary server. The remaining network interfaces are backups; the servers providing these interfaces are the backup servers. By default, HP Clustered File System directs network traffic for the virtual host to the primary interface.

The primary and backup servers do not need to be dedicated to these activities; all servers can support other independent functions.

The network interface currently receiving the network traffic is the *active interface* and is labeled as "Active" on the Management Console. This interface is typically the primary interface; however, if HP Clustered File System has redirected the network traffic to a backup interface, that interface becomes the active interface. Network interfaces not currently receiving network traffic for a virtual host are called *inactive interfaces*.

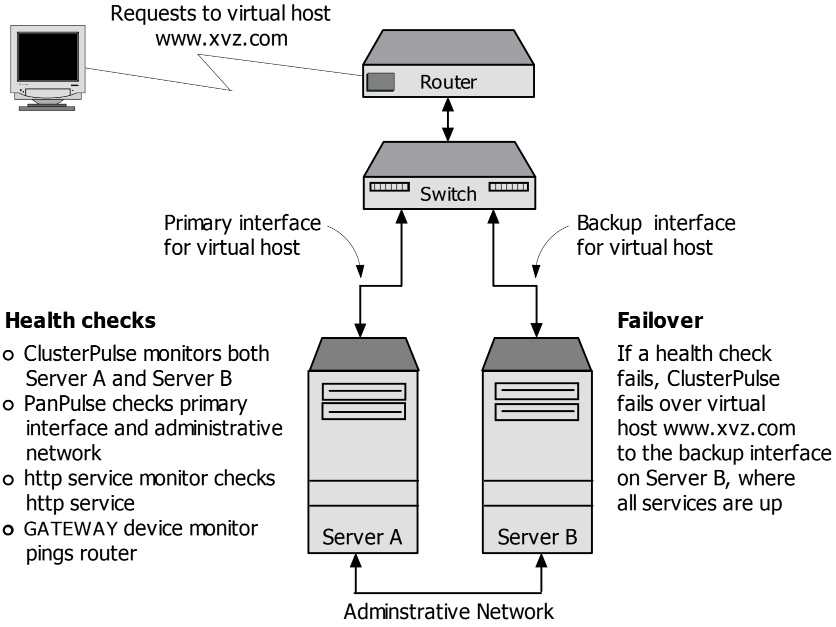
After creating virtual hosts, you will need to configure your network applications to recognize them. When clients want to access a network application, they use the virtual host address instead of the address of the server where the application is running.

Cluster Health and Virtual Host Failover

To ensure the availability of a virtual host, HP Clustered File System monitors the health of the administrative network, the active network interface, and the underlying server. If you have created service or device monitors, those monitors periodically check the health of the specified services or devices.

If any of these checks fail, HP Clustered File System can transfer the virtual host to a backup interface on another server and the network traffic will continue.

The following example shows the virtual host failover mechanism. Servers A and B provide the http service, which is accessed with the virtual host www.xvz.com. Each server is configured with an http service monitor and a gateway device monitor to ping the router.


Requests to virtual host
www.xvz.com
Router
Switch
Primary interface
for virtual host
Backup interface
for virtual host
Health checks
o ClusterPulse monitors both Server A and Server B
o PanPulse checks primary interface and administrative network
o http service monitor checks http service
o GATEWAY device monitor pings router
Failover
If a health check fails, ClusterPulse fails over virtual host www.xvz.com to the backup interface on Server B, where all services are up
Server A
Server B
Adminstrative Network

The failover operation to another network interface has minimal impact on clients. For example, if clients were downloading Web pages during the failover, they would receive a "transfer interrupted" message and could simply reload the Web page. If they were reading Web pages, they would not notice any interruption.

If the active network interface fails, only the virtual hosts associated with that interface are failed over. If the underlying server has other virtual hosts configured on another network interface, those virtual hosts will continue to access that server.

For more information about failover, see "Virtual Hosts and Failover" on page 119.

Guidelines for Creating Virtual Hosts

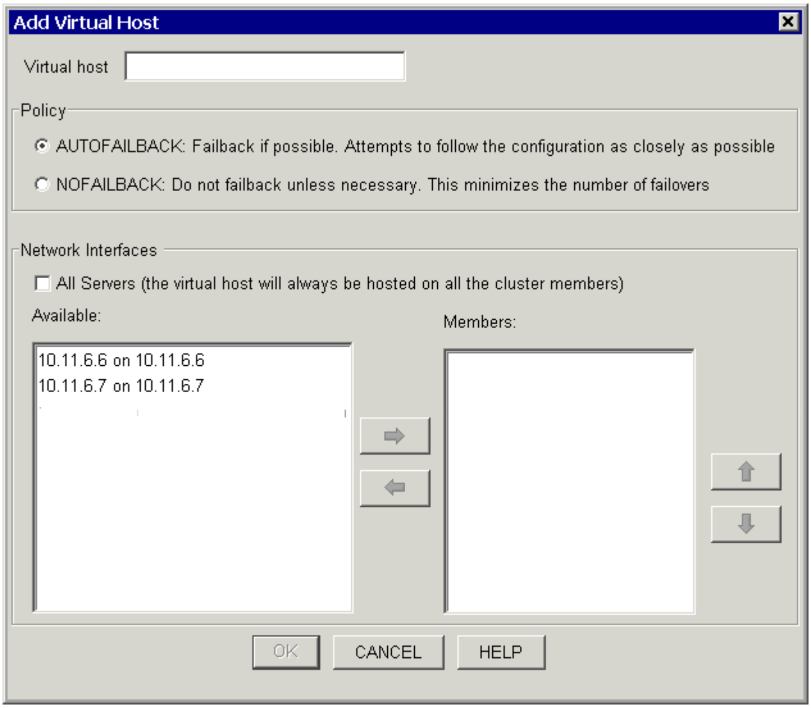
When creating virtual hosts, follow these guidelines:

- When planning the virtual hosts needed for your cluster, first determine the network services that will be available to your clients. Then determine the IP addresses for those services. You will need to create a virtual host for each IP address.
- Choose virtual hostnames that differ from your actual server names. Virtual hosts are independent of specific servers, and their names should be independent as well.
- Use an IP address that is on the same subnet as the network interfaces where it will be configured.
- Update the DNS name service or the */etc/hosts* file with the virtual hostnames and IP addresses. (For improved performance, the Management Console caches hostname lookups. If your DNS changes, you may need to restart the console so that it will reflect the new hostname.)
- Do not use **ifconfig** or another tool to configure the IP address in the operating system or a system startup script. HP Clustered File System configures the operating system appropriately to support the virtual host.
- After creating virtual hosts, you will need to configure your applications to recognize them.

Add or Modify a Virtual Host

To add or update a virtual host, select the appropriate option:

- To add a new virtual host, select Cluster > Virtual Host > New or click the Add Virtual Host icon on the toolbar. Then configure the virtual host on the Add Virtual Host window.
- To update an existing virtual host, select that virtual host on either the Server or Virtual Hosts window, right-click, and select Properties. Then make your changes on the Virtual Host Properties window.



Add Virtual Host

Virtual host

Policy

● AUTOFAILBACK: Failback if possible. Attempts to follow the configuration as closely as possible

○ NOFAILBACK: Do not failback unless necessary. This minimizes the number of failovers

Network Interfaces

□ All Servers (the virtual host will always be hosted on all the cluster members)

Available:

10.11.6.6 on 10.11.6.6
10.11.6.7 on 10.11.6.7

Members:

Virtual Host: Type a hostname or an IP address for the virtual host.

Policy: The policy determines the failback action that the virtual host will take following a failover to a backup server.

As part of configuring a virtual host, you will need to select network interfaces located on the servers that can be used for the virtual host.

The interfaces are placed in order: primary, backup #1, backup #2, and so on. The ClusterPulse process considers the "health" of the servers providing those interfaces when determining where to place a virtual host. The status and enablement of each monitor contributes to the health of a server with respect to the virtual host. When a server is completely "healthy," all of the services associated with the virtual host are up and enabled.

- AUTOFAILBACK. This is the default policy. It is intended to return the virtual host to its original configuration, or as close to it as possible. After the virtual host fails over to a backup server, HP Clustered File System watches the health of the servers higher in the list for that virtual host. When the health of one of these servers is equal to or greater than the backup server where the virtual host currently resides, the virtual host will automatically attempt to fail over to that server.
- NOFAILBACK. This policy is intended to minimize failovers. The virtual host remains active on the backup server until a "healthier" server becomes available, at which point the virtual host fails over to that server. (On a "healthier" server, more of the services associated with the virtual host will be up than on the server currently hosting the virtual host.)

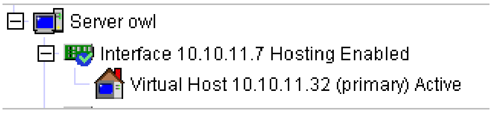
Network Interfaces: When the "All Servers" box is checked, the virtual host will be configured on all servers having an interface on the network you select for this virtual host. When you add another server to the cluster, the virtual host will automatically be configured on that server. This option can be useful with administrative applications.

Available:/Members: The Available column lists all network interfaces that are available for this virtual host. If you are configuring the virtual host on all servers in the cluster, move the network interface for the primary server to the Members column. The corresponding network interfaces on the remaining servers will then be moved to the Members column automatically.

If you are configuring the virtual host only on certain servers, move the appropriate network interfaces for those servers to the Members column.

The first interface you select will be the primary interface. The other interfaces you select will be backups.

The virtual host now appears on both the Virtual Hosts window and the Servers window, as shown below.



Server owl
Interface 10.10.11.7 Hosting Enabled
Virtual Host 10.10.11.32 (primary) Active

To add or update a virtual host from the command line, use the following command. The first network interface specified is the primary interface; the remaining interfaces are backups.

```
mx vhost add|update <vhost> <network_interface> ...
```

Configure Applications for Virtual Hosts

After creating virtual hosts, you will need to configure your network applications to recognize them. For example, if you are using a Web server, you may need to edit its configuration files to recognize and respond to the virtual hosts. By default, FTP responds to any virtual host request it receives.

You do not need to configure operating system software or files to respond to the virtual host; HP Clustered File System does that configuration for you. For example, do not edit start-up scripts or run the **ifconfig** command with the virtual host information.

Other Virtual Host Procedures

Delete a Virtual Host

Select the virtual host to be deleted on either the Servers window or the Virtual Hosts window, right-click, and select Delete. Any service monitors configured on that virtual host are also deleted.

To delete a virtual host from the command line, use this command:

```
mx vhost delete <vhost> ...
```

Change the Virtual IP Address for a Virtual Host

When you change the virtual IP address of a virtual host, you will also need to update your name server and to configure applications to recognize the new virtual IP address. The order in which you perform these tasks is dependent on your application and the requirements of your site.

You can use **mx** commands to change the virtual IP address of a virtual host. Complete these steps:

1. At the command prompt, run the following command:

   ```
   mx matrix dump > mxdump.out
   ```

2. Edit the *mxdump.out* file, replacing all occurrences of the old IP address with the new IP address.
3. At the beginning of the *mxdump.out* file, add a line such as the following to remove the old IP address from the virtual host configuration:

   ```
   vhost delete <old_IP_address>
   ```

4. Delete any lines in *mxdump.out* that do not need to change.
5. Run the following command to update the cluster configuration with the new IP address:

   ```
   mx --file mxdump.out
   ```

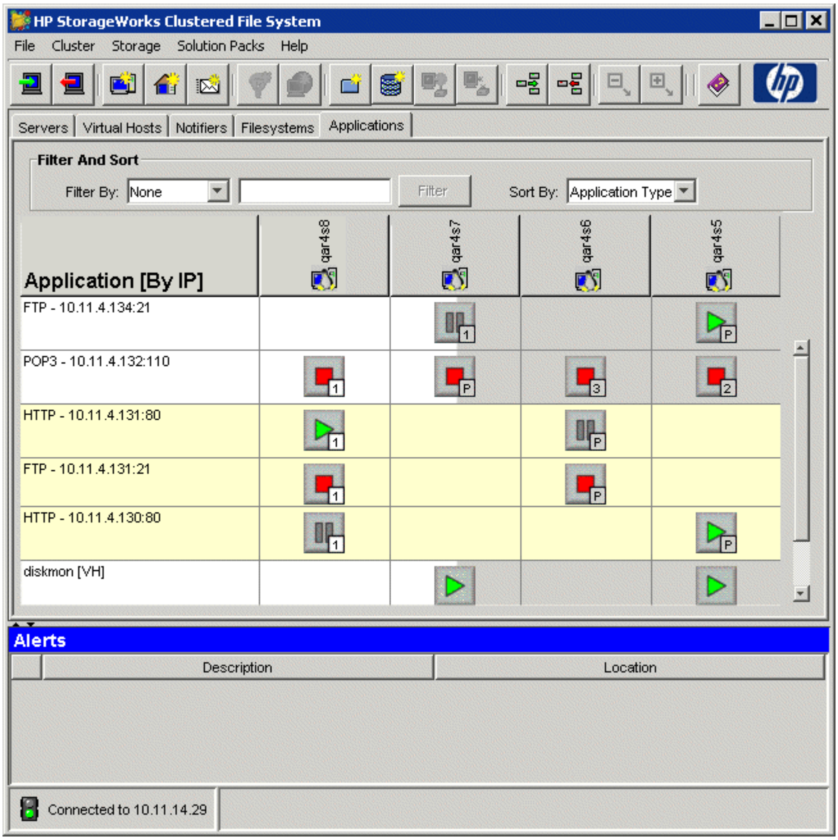
Using the Applications Tab

The Applications tab on the Management Console shows all application monitors running in the cluster and enables you to manage and monitor them from a single screen.

The Applications tab displays the cluster servers and application monitors in a table format. The servers appear in the columns across the top of the tab. The rows of the table list the application monitors. Double-clicking on an application monitor brings up the properties for that monitor.

Monitors are displayed for two types of applications:

- **HA applications.** These monitors include the built-in and custom service monitors provided with HP Clustered File System. The monitors are grouped according to the virtual hosts on which they are deployed.
- **Scale-out applications.** These applications include the built-in and custom device monitors provided with HP Clustered File System. These monitors are deployed on the servers in the cluster and appear after the HA applications on the Applications tab.



Filter By: None

Sort By: Application Type

Application [By IP]	qar4s8	qar4s7	qar4s6	qar4s5
FTP - 10.11.4.134:21		⏸ 1		▶ P
POP3 - 10.11.4.132:110	🟥 1	🟥 P	🟥 3	🟥 2
HTTP - 10.11.4.131:80	▶ 1		⏸ P	
FTP - 10.11.4.131:21	🟥 1		🟥 P	
HTTP - 10.11.4.130:80	⏸ 1			▶ P
diskmon [VH]		▶		▶

	Description	Location

Connected to 10.11.14.29

If only one application monitor is configured on a virtual host, the row for that application is colored white; if there are multiple application monitors on a virtual host, the rows are yellow. The scale-out applications appear in white following the HA applications (diskmon in the above example).

The Applications tab shows whether a monitor is deployed on a particular server, as well as the current status of the monitor. If a cell is empty, the monitor is not deployed on that server.

The icons used on the Applications tab have the following meanings:

▶P The green arrow specifies that the HA application monitor is active on this server. In this example, “P” indicates that this is the primary interface.

❚❚1 The grey bars indicate that the HA application monitor is inactive on this server. In this example, “1” indicates that this is the first backup.

❚❚2 The red bars indicate that the HA application monitor is disabled on this server. In this example, “2” indicates that this is the second backup.

❚❚2 The yellow bars indicate that the HA application monitor is transitioning to another state.

■1 The red square specifies that the HA application monitor is down.

▶ The green arrow specifies that the scale-out application monitor is up. This status applies to the entire cluster.

▶ The red arrow specifies that the scale-out application monitor is down. This status applies to the entire cluster.

Modify the Display

You can change the order of the server columns by dragging a column to another location. You can also resize the columns.

You can also limit the information appearing on the Application tab. For example, you may want to see only a certain type of application monitor, or only application monitors that are down or disabled. You can use the Filters And Sorts options to do this.

Sort Options

By default, applications on a particular virtual host are listed in descending order according to the IP address assigned to the application. You can reorder the application monitors to be in descending order by name by using the Application Type sort option. The IP sort option restores the default order.

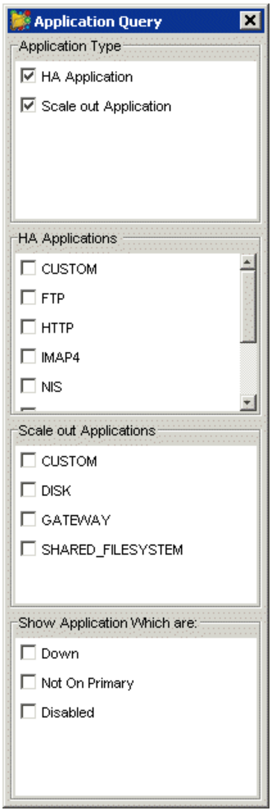
Filter Options

The filter options are as follows:

None. No filtering takes place.

Virtual Host. This option allows you to filter by a specific virtual host. To use this option, enter the hostname or IP address of the virtual host and then click the Filter button.

Options. This selection brings up the Application Query dialog, which allows you to limit the information that appears on the Applications tab.



Application Query

Application Type

☑ HA Application

☑ Scale out Application

HA Applications

□ CUSTOM

□ FTP

□ HTTP

□ IMAP4

□ NIS

Scale out Applications

□ CUSTOM

□ DISK

□ GATEWAY

□ SHARED_FILESYSTEM

Show Application Which are:

□ Down

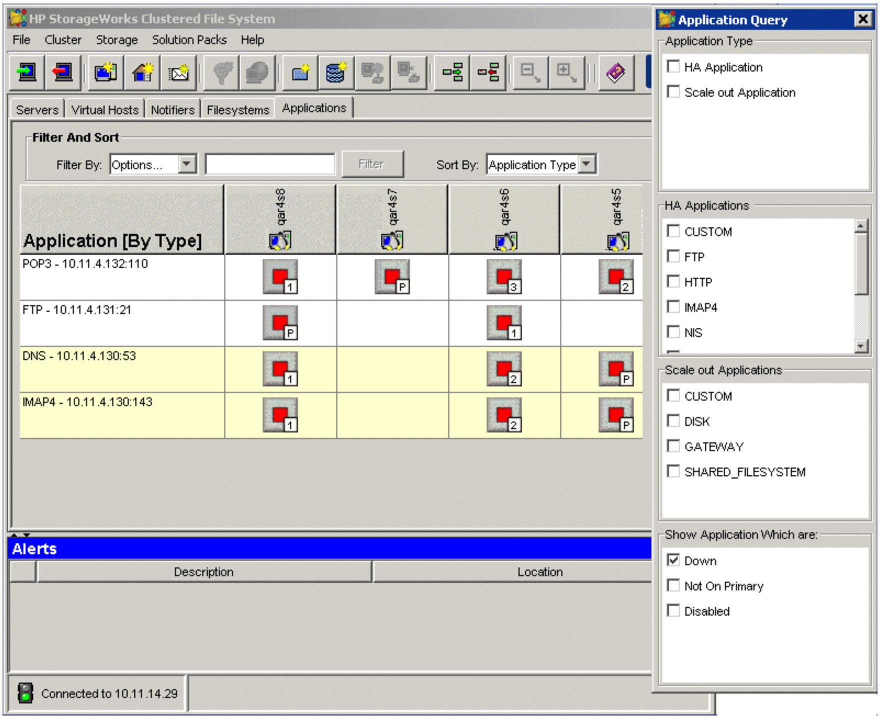
□ Not On Primary

□ Disabled

By default, HA Application and Scale-out Application are checked at the top of the dialog. These selections cause all of the information for HA and scale-out applications to be displayed. To limit the information, uncheck either or both of these items. Then select the types of application monitors or monitor status that you want to see. The display immediately updates to reflect your selections.

NOTE: The display reflects the union of all checked filters. For example, if both HA Application and Disabled are checked, the display will include all HA Application monitors and all Disabled monitors.

The following example shows only the application monitors that are down.



Filter By: Options...

Sort By: Application Type

Application [By Type]	qar4s8	qar4s7	qar4s6	qar4s5
POP3 - 10.11.4.132:110	🟥 1	🟥 P	🟥 3	🟥 2
FTP - 10.11.4.131:21	🟥 P		🟥 1	
DNS - 10.11.4.130:53	🟥 1		🟥 2	🟥 P
IMAP4 - 10.11.4.130:143	🟥 1		🟥 2	🟥 P

	Description	Location

Connected to 10.11.14.29

Application Type
- □ HA Application
- □ Scale out Application

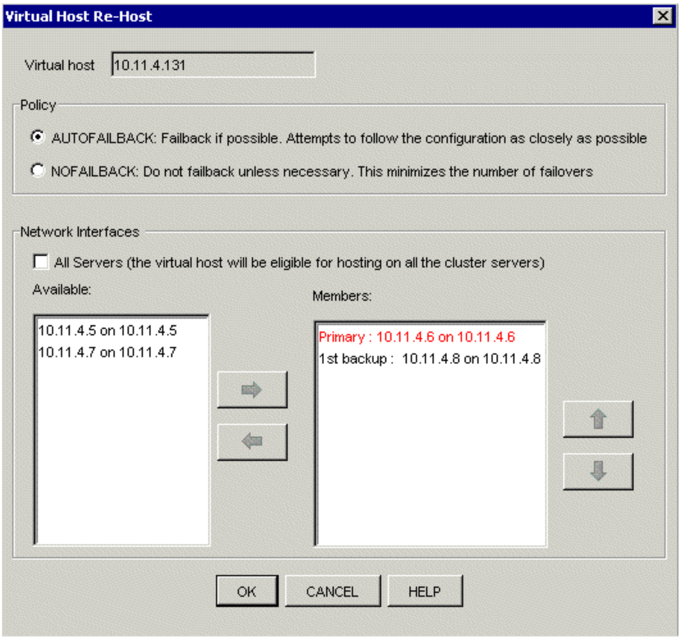
HA Applications
- □ CUSTOM
- □ FTP
- □ HTTP
- □ IMAP4
- □ NIS

Scale out Applications
- □ CUSTOM
- □ DISK
- □ GATEWAY
- □ SHARED_FILESYSTEM

Show Application Which are:
- ☑ Down
- □ Not On Primary
- □ Disabled

Re-Host Virtual Hosts

You can use the Applications tab to modify the configuration of a virtual host. For example, you might want to change the primary for the virtual host. To re-host a virtual host, right click in a cell for that virtual host and then select Re-Host. You will then see a message warning that this action will cause the clients of the application to lose connection. When you continue, the Virtual Host Re-Host window appears.



Virtual Host Re-Host

Virtual host: 10.11.4.131

Policy

● AUTOFAILBACK: Failback if possible. Attempts to follow the configuration as closely as possible

○ NOFAILBACK: Do not failback unless necessary. This minimizes the number of failovers

Network Interfaces

□ All Servers (the virtual host will be eligible for hosting on all the cluster servers)

Available:

10.11.4.5 on 10.11.4.5
10.11.4.7 on 10.11.4.7

Members:

Primary : 10.11.4.6 on 10.11.4.6
1st backup : 10.11.4.8 on 10.11.4.8

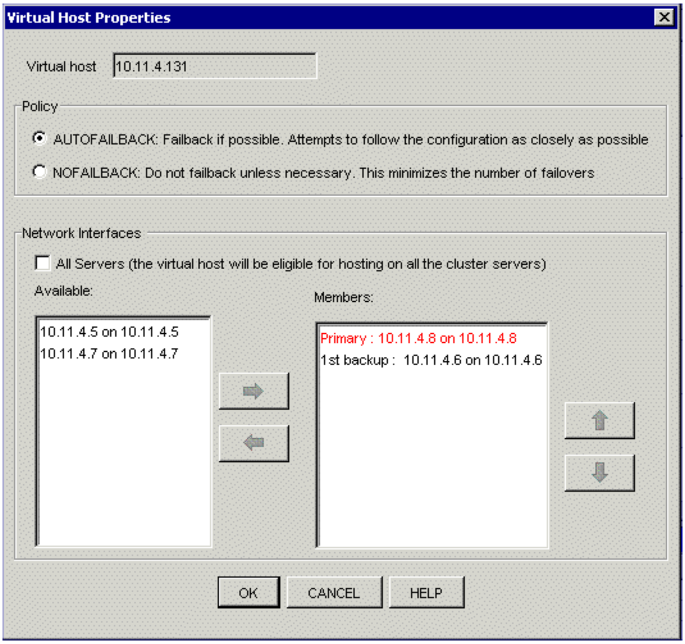
You can now make the necessary changes to the virtual host.

Manage Applications

The Applications tab provides several ways to manage applications and virtual hosts.

Add or Remove a Server

HA Applications: When you click on Add Server to Application, the Virtual Host Properties window appears. Move the network interface for the server to the Members column and click OK.



Virtual Host Properties

Virtual host 10.11.4.131

Policy

● AUTOFAILBACK: Failback if possible. Attempts to follow the configuration as closely as possible

○ NOFAILBACK: Do not failback unless necessary. This minimizes the number of failovers

Network Interfaces

□ All Servers (the virtual host will be eligible for hosting on all the cluster servers)

Available:

10.11.4.5 on 10.11.4.5
10.11.4.7 on 10.11.4.7

Members:

Primary : 10.11.4.8 on 10.11.4.8
1st backup : 10.11.4.6 on 10.11.4.6

This step adds the server to the virtual host associated with the HA application monitor. All application monitors configured on that virtual host will be added to the server.

To remove an application monitor from a server, right-click in the cell for that monitor/server combination, select Remove From Server, and verify that you want to remove the monitor. This step removes the network interface for the server from the configuration of the virtual host associated with the monitor. Because the server is no longer part of the virtual host configuration, all application monitors on that virtual host are removed from the server.

Scale out applications: Click in the empty cell corresponding to the application and server, right-click, and select Add Server to Application. The application monitor is then added to the server.

To remove a scale-out application monitor from a server, click on the cell containing the monitor, right-click, and select Remove From Server.

Other Application Management Procedures

For HA application monitors, you can perform the same procedures as are available when right-clicking on a service monitor on the Servers or Virtual Hosts tab. For scale-out application monitors, you can perform the procedures available when right-clicking on a device monitor on the Servers tab.

The procedures are as follows:

- Delete the application monitor
- Modify the configuration of the application monitor
- Enable or disable the application monitor on a specific server
- View or clear the last error

For more information about using these procedures on HA application monitors, see "Add or Modify a Service Monitor" on page 129 and "Other Configuration Procedures" on page 137.

For more information about using these procedures on scale-out application monitors, see "Add or Modify a Device Monitor" on page 144 and "Other Configuration Procedures" on page 153.

Virtual Hosts and Failover

When you create a virtual host, you specify a list of network interfaces on which the virtual host can be located. The interfaces are placed in order: primary, backup #1, backup #2, and so on. The ClusterPulse process considers the "health" of the servers providing those interfaces when determining where to place a virtual host. The status and enablement of the service and device monitors associated with the virtual host also contribute to a server's health calculation. When a server is completely "healthy," all of the services associated with the virtual host are up and enabled.

When certain events occur on the server where a virtual host is located, the ClusterPulse process will attempt to fail over the virtual host to another server configured for that virtual host. For example, if the server goes down, ClusterPulse will check the health of the other servers and then determine the best location for the virtual host.

ClusterPulse uses the following virtual-host activeness policy to determine the server where it will make a virtual host active. In conjunction with this policy, the decisions that you make when configuring a virtual host and the service or device monitors associated with it help determine whether virtual host failover occurs, the interface to which the virtual host will fail over, and what happens when operations are restored on the original server.

Virtual Host Activeness Policy

The policy described here is accurate for this release but it may change in future releases.

The virtual host activeness policy decision is made as follows:

1. If the virtual host is disabled, it is not made active anywhere.
2. ClusterPulse considers the list of servers that are both up and enabled and that are configured for the virtual host. The network interface that the virtual host is associated with must also be both up and enabled for hosting. Note the following:

    - A server that has not finished joining the cluster (see "Server Access to the SAN" on page 179) is not considered up for the purpose of activating the device monitor.
    - A server is considered down if it loses coordinated communication with the cluster (for example, the server crashed or was shut down, HP Clustered File System was shut down on that server, the server failed to schedule a cluster group communication process for an extended period of time, the server disabled the NIC being used for cluster network traffic, and so on).
    - The PanPulse process controls whether a network interface is marked up or down. When PanPulse determines that an interface currently hosting a virtual host is down, ClusterPulse will begin searching for another server on which to locate the virtual host.

3. ClusterPulse narrows the list to those servers without inactive, down, or disabled HP Clustered File System device monitors. If there are no servers that meet this criteria, the virtual host is not made active anywhere.
4. If the virtual host is currently active on a server and that virtual host has the NOFAILBACK policy, then this active server is moved to the head of the list of preferred servers to be considered.
5. From this list, ClusterPulse tries to find a server with all services up and enabled. If ClusterPulse finds a server meeting these conditions, it will use it, preferring servers earlier in the list of servers configured for the virtual host.
6. If there are no servers with completely healthy services, ClusterPulse picks a server that has at least one service up and enabled. If the services have different priorities, ClusterPulse uses a server where a service with the highest available priority is up. If ClusterPulse finds a server meeting these conditions, it will use it, preferring servers earlier in the list of servers configured for the virtual host.
7. The selected server will have one interface that was configured for this virtual host. The virtual host will be active on this interface. If ClusterPulse cannot locate a server meeting these conditions, it does not place the virtual host anywhere.

Customize Service and Device Monitors for Failover

By default, when a service or device monitor probe fails, indicating that the watched service is down or the monitored device cannot be accessed, ClusterPulse will fail over the associated virtual host to another server where the monitored service or device is up.

You can customize this behavior using the Advanced monitor settings.

Specify Failover/Failback Behavior

The Probe Severity setting allows you to specify whether a failure of the service or device monitor probe should cause the virtual host to fail over. For example, you could configure a gateway device monitor to watch a router. The device monitor probe might occasionally time out because of heavy network traffic to the router; however the router is still functioning. In this case, you can configure the device monitor so that the virtual host will not fail over if only the device monitor has failed.

The Probe Severity setting also affects the failback behavior of the virtual host.

There are three settings for Probe Severity:

- NOFAILOVER. The virtual host does not fail over when a monitor probe fails.
- AUTORECOVER. The virtual host fails over when a monitor probe fails. When the service or device is recovered on the original node, failback occurs according to the virtual host's failback policy.
- NOAUTORECOVER. The virtual host fails over when a monitor probe fails and the monitor is disabled on the original node, preventing automatic failback. When the monitor is reenabled, failback occurs according to the virtual host's failback policy

Select a Backup Interface for Failover

When ClusterPulse needs to fail over a virtual host, it determines the best network interface to receive the virtual host. It bases its decision on the status of the available backup interfaces and their underlying servers. Typically, it looks for an "up" network interface where all services are "up" on the underlying server.

You can use the following Advanced settings to affect how ClusterPulse selects the network interface for failover.

- The Event Severity setting allows you to specify whether ClusterPulse should consider the existence of monitor events (such as a script failure or timeout) when it chooses a network interface for failover. If the events are considered, the network interface for the affected server becomes less desirable.
- For service monitors, you can assign a priority to each monitor (the Service Priority setting). If ClusterPulse cannot locate an interface where all services are "up" on the underlying server, it selects an interface where the highest priority service is "up" on the underlying server.

Specify Failback Behavior of the Virtual Host

When you configure a virtual host, you can specify the failback action that the virtual host will take following a failover to a backup server. There are two options: AUTOFAILBACK and NOFAILBACK.

AUTOFAILBACK is the default for virtual hosts.

AUTOFAILBACK

AUTOFAILBACK is intended to return the virtual host to its original configuration, or as close to it as possible. After the virtual host fails over to a backup server, ClusterPulse watches the health of the servers higher in the list for that virtual host. When the health of one of these servers is *equal to or greater* than the backup server where the virtual host currently resides, the virtual host automatically attempts to fail over to that server.

NOFAILBACK

NOFAILBACK is intended to minimize failovers. The virtual host remains active on the backup server until a healthier server (it's health is *greater* than the backup server) becomes available, at which point the virtual host will fail over to that server.

For example, consider a two-node cluster that is configured with its primary on node 1 and backup on node 2, and that uses the NOFAILBACK option. Three service monitors are configured on the virtual host. When a service monitor probe fails on node 1, the virtual host will fail over to node 2. Following are some possible failback scenarios:

- When the monitored service is restored on node 1, the virtual host will remain on node 2. Node 1 and node 2 are equally healthy; they both have three up service monitors. The NOFAILBACK policy will not move the virtual host until a *healthier* server is available.
- After the virtual host fails over to node 2, a service monitor probe fails on that node. Now both nodes have a down service monitor. Failback does not occur because the servers are equally healthy. If the failed service is then restored on node 1, that node will now be healthier than node 2 and failback will occur. (Note that if the virtual host policy was AUTOFAILBACK, failback would occur when the probe failed on node 2 because both servers were equally healthy.)

- After the virtual host fails over to node 2, all service monitor probes fail on that node. Node 1, with one down monitor, is now healthier than node 2, with three down monitors, and failback will occur.

Failback Policy and Monitor Probe Severity

The following table shows how the virtual host failback policy interacts with the probe severity setting for service and device monitors.

Virtual Host Policy	Monitor Probe Severity	Behavior When Probe Reports DOWN
AUTOFAILBACK	NOFAILOVER	Failover does not occur.
	AUTORECOVER	Failover occurs. When service is restored, failback occurs.
	NOAUTORECOVER	Failover occurs and the monitor is disabled on the original server. When the monitor is reenabled, failback occurs.
NOFAILBACK	NOFAILOVER	Failover does not occur.
	AUTORECOVER	Failover occurs. The virtual host remains on the backup server until a “healthier” server is available.
	NOAUTORECOVER	Failover occurs and monitor is disabled on the original server. The virtual host remains on the backup server until a “healthier” server is available.

10

Configure Service Monitors

Service monitors are typically used to monitor a network service such as HTTP or FTP. If a service monitor indicates that a network service is not functioning properly on the primary server, HP Clustered File System can transfer the network traffic to a backup server that also provides that network service.

Overview

Before creating a service monitor for a particular service, you will need to configure that service on your servers. For example, you could configure Netscape or Apache to provide HTTP service and then create a service monitor for HTTP. A service can run regardless of HP Clustered File System monitoring.

Service Monitors and Virtual Hosts

A service monitor is associated with a specific virtual host; the monitored service must be configured on all servers supporting that virtual host. Using virtual hosts enables HP Clustered File System to switch traffic quickly between servers. If the service being monitored goes down on the primary server, HP Clustered File System moves the virtual host traffic to a backup server, which then handles the traffic for the service.

If you add another server to the configuration for a virtual host, the service monitors assigned to that virtual host are automatically added to the new server.

Similarly, if a server is removed from the configuration, the service monitors assigned to the virtual host are automatically removed from that server. Service monitor parameters (such as probe severity, Start scripts, and Stop scripts) are consistent across all servers configured for a virtual host.

Service Monitors and Failover

If a monitored service fails, HP Clustered File System attempts to relocate any virtual hosts associated with the service monitor to a network interface on a healthier server. However, if there are multiple failures in the cluster or the virtual host does not have a backup network interface on another server, HP Clustered File System may not be able to move the virtual host. The virtual host may then remain active on a server with a failed monitored service. HP Clustered File System will avoid this situation if there is a healthy server configured and available, but sometimes it may not be possible in a multiple-failure scenario.

Types of Service Monitors

HP Clustered File System supports the following service monitors. When configuring a monitor, you will need to supply the following information.

Type	Port	Default Probe Timeout	Default Probe Frequency	Script Parameters
FTP	21	5 seconds	30 seconds	none
HTTP	80	5 seconds	30 seconds	URL
IMAP4	143	5 seconds	30 seconds	none
NFS	2049	5 seconds	30 seconds	protocol, rpcname, rpcvers
NIS	0	5 seconds	30 seconds	protocol, rpcname, rpcvers, domain
NNTP	119	5 seconds	30 seconds	none
POP3	110	5 seconds	30 seconds	none
SMTP	25	5 seconds	30 seconds	none

Type	Port	Default Probe Timeout	Default Probe Frequency	Script Parameters
TCP	0	5 seconds	30 seconds	none
CUSTOM	NA	60 seconds	60 seconds	user probe script

FTP Service Monitor

By default the FTP service monitor probes TCP port 21 of the virtual host address. You can change this port number to the port number configured for your FTP server. The default frequency of the probe is every 30 seconds. The default time that the service monitor waits for a probe to complete is five seconds.

The probe function attempts to connect to port 21 and expects to read an initial message from the FTP server. The message should begin with 220. If the probe is successful, the monitor sends the string QUIT\n to the server. If there are no errors, the service status remains Up. If an error occurs, the status is set to Down.

HTTP Service Monitor

By default, the HTTP service monitor probes TCP port 80 of the virtual host address. You can change this port number to the port number configured for your HTTP server. The default frequency for the probe is every 30 seconds; the default time that the service monitor waits for a probe to complete is five seconds.

You can also supply a URL (either a directory or a full URL) that will be read as part of the probe function. The probe attempts to connect to the HTTP server and read the URL. If the response from the HTTP server is a success code (200 <= reply < 300), then the service status remains Up. If the probe does not receive a success code or it takes more than five seconds to get the reply from the HTTP server, the service status is set to Down. (You can view the reply code from the Management Console; select the service monitor, right-click, and select View Last Error.)

NFS Service Monitor

The NFS service monitor only monitors the health of an NFS server. For high availability and failover support, use the MxFS-Linux Solution Pack.

SMTP Service Monitor

By default, the SMTP service monitor probes TCP port 25 (sendmail port) of the virtual host address. You can change this port number to the port number configured for your SMTP server. The default frequency of the probe is every 30 seconds. The default time that the service monitor waits for a probe to complete is five seconds.

The probe attempts to connect to port 25 and expects to read an initial message from the SMTP server. This message should begin with 220. If the probe is successful, the monitor sends the string QUIT\n to the SMTP server. If there are no errors, the service status remains Up. If an error occurs, the status is set to Down.

TCP Service Monitor

The generic TCP service monitor defaults to TCP port 0. You should set the port to the listening port of your server software. The default frequency of the probe is every 30 seconds. The default time that the service monitor waits for a probe to complete is five seconds. Because the service monitor cannot know what to expect from the TCP port connection, it simply attempts to connect to the specified port. If the attempt is successful, the service status remains Up.

NOTE: Service monitors using TCP port numbers must be assigned distinct port numbers. For example, the built-in FTP service monitor and a generic TCP monitor cannot both attempt to use TCP port 21. You do not need to specify an IP address because a service monitor is associated with a virtual host. HP Clustered File System uses that virtual host IP address in combination with the TCP port number to create network TCP connections for service monitors.

Custom Service Monitor

HP Clustered File System provides a CUSTOM service monitor type that can be used when the built-in monitor types are not sufficient. Custom monitors can be particularly useful when integrating HP Clustered File System with a custom application.

HP Clustered File System treats custom monitors just as it does the built-in monitors, except that you must supply the probe script. In the script, probe commands should determine the health of the service as necessary. If the service is operating normally, the probe script should exit with exit status zero. If the service is not operating normally, the probe script should exit with a non-zero exit status. HP Clustered File System will interpret the non-zero exit status as a failure of the service and then take the appropriate action.

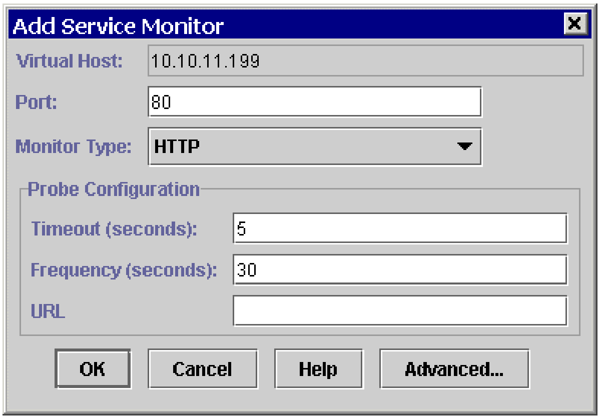
See "Advanced Topics" on page 166 for information about developing probe scripts for custom monitors and integrating monitors with custom applications.

Add or Modify a Service Monitor

Adding a service monitor configures HP Clustered File System monitoring only. It does not configure the service itself.

To add or update a service monitor, select the appropriate option:

- To add a new service monitor, first select the virtual host for the monitor on either the Servers or Virtual Hosts window, then right-click and select Add Service Monitor (or click the Add Service Monitor icon on the toolbar). Then configure the monitor on the Add Service Monitor window.
- To modify an existing service monitor, select that monitor on either the Servers or Virtual Hosts window, right-click, and select Properties. Then make your changes on the Service Monitor Properties window.



Add Service Monitor

Virtual Host: 10.10.11.199

Port: 80

Monitor Type: HTTP

Probe Configuration

Timeout (seconds): 5

Frequency (seconds): 30

URL

Virtual Host: The service monitor is assigned to this virtual host.

Port: HP Clustered File System supplies the default port number for the service you select. If your service uses a port other than the default, type that port number here.

Monitor Type: Select the type of service that you want to monitor.

Timeout: The maximum amount of time that the **monitor_agent** daemon will wait for a probe to complete. For most monitors, the default timeout interval is five seconds. You can use the default setting or specify a new timeout interval.

Frequency: The interval of time, in seconds, at which the monitor probes the designated service. You can use the default setting, typically 30 seconds, or enter a new frequency interval.

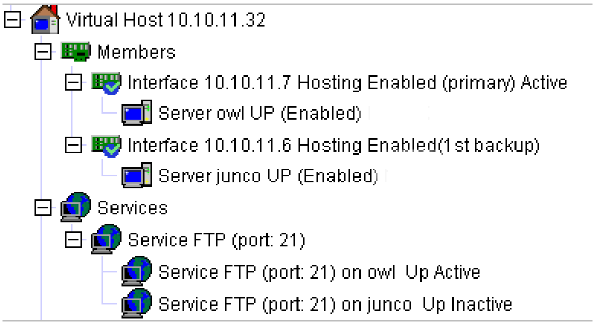
Script parameters: If you are configuring a HTTP, NFS, NIS, or CUSTOM monitor, this field will ask for the following information. (This field is unused for the other types of monitors.)

- **HTTP service monitor.** You will be asked to specify the URL for the HTTP service. When the monitor performs its probe operation, it will attempt to retrieve the URL from the server, giving a higher level of assurance that the service is functioning correctly. If you do not specify a URL, the probe operation will connect to **httpd** and wait for a response. If **httpd** responds, the monitor assumes that the service is operating correctly.

- **NFS service monitor.** You will be asked to set the **proto**, **rpcname**, and **rpcvers** parameters. Set **proto** to either **udp** or **tcp**, depending on the protocol configured for NFS. **rpcname** is typically set to **nfs**. Set **rpcvers** to either **2** or **3** to match the protocol of your NFS configuration (either version 2 or version 3 protocol).
- **NIS service monitor.** You will be asked to set the **proto**, **rpcname**, **rpcvers**, and **domain** parameters. Set **proto** to either **udp** or **tcp**, depending on the protocol used by the NIS protocol version you have configured. **rpcname** is typically set to **ypserv**. Set **rpcvers** to either **1** or **2** to match the NIS protocol version used by your server. Set **domain** to the NIS domain that should be monitored by the NIS probe.

  The **rpcinfo** command can be useful for determining the correct values for **proto**, **rpcname**, and **rpcvers**.
- **Custom service monitor.** You will be asked for the location of the user probe script. Enter the pathname for the probe script be used with the monitor. See “Custom Scripts” on page 166 for information about writing probe scripts.

When you complete the Add Service Monitor form, the new monitor appears on the Management Console. In this example, the service monitor is active on server owl, which provides the active network interface for the virtual host.



```
⊟ Virtual Host 10.10.11.32
  ⊟ Members
    ⊟ Interface 10.10.11.7 Hosting Enabled (primary) Active
        Server owl UP (Enabled)
    ⊟ Interface 10.10.11.6 Hosting Enabled(1st backup)
        Server junco UP (Enabled)
  ⊟ Services
    ⊟ Service FTP (port: 21)
        Service FTP (port: 21) on owl  Up Active
        Service FTP (port: 21) on junco  Up Inactive
```

To add or update a service monitor from the command line, use this command:

```
mx service add|update [--type <type>] [--timeout <seconds>]
[--frequency <seconds>] [<other_arguments>] <vhost:port> ...
```

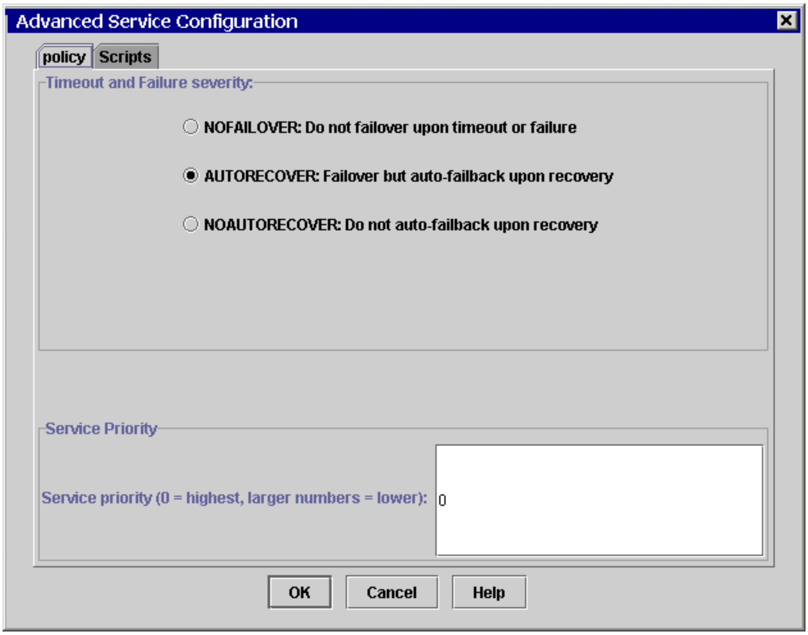
See "Advanced Settings for Service Monitors" for information about the other arguments that can be specified for service monitors.

Advanced Settings for Service Monitors

You can use the advanced settings to fine-tune service monitors. To access the advanced settings, select the Advanced button on the Add Service Monitor or Service Monitor Properties window.

Service Monitor Policy

The Policy tab lets you specify the failover behavior of the service monitor and set its service priority.


Advanced Service Configuration

policy | Scripts

Timeout and Failure severity:

○ NOFAILOVER: Do not failover upon timeout or failure

● AUTORECOVER: Failover but auto-failback upon recovery

○ NOAUTORECOVER: Do not auto-failback upon recovery

Service Priority

Service priority (0 = highest, larger numbers = lower): 0

Timeout and Failure Severity

This setting works with the virtual host policy (either AUTOFAILBACK or NOFAILBACK) to determine what happens when a probe of a monitored service fails.

The default policies (AUTOFAILBACK for the virtual host and AUTORECOVERY for the monitor) cause ClusterPulse to fail over the associated virtual host to a backup network interface on another server. When the service is recovered on the original node, the virtual host fails back to that node.

You can use the Timeout and Failure Severity attribute to change this behavior. There are three settings: NOFAILOVER, AUTORECOVER, and NOAUTORECOVER.

NOFAILOVER. When the monitored service fails, ClusterPulse does not fail over to a backup network interface. This option is useful when the monitored resource is not critical, but is important enough that you want to keep a record of its health.

AUTORECOVER. This is the default. The virtual host fails over when a monitor probe fails. When the service is recovered on the original node, failback occurs according to the virtual host's failback policy.

NOAUTORECOVER. The virtual host fails over when a monitor probe fails and the monitor is disabled on the original node, preventing automatic failback. When the monitor is reenabled, failback occurs according to the virtual host's failback policy.

The NOAUTORECOVER option is useful when integrating HP Clustered File System with a custom application where certain application-specific actions must be taken before the failback can occur.

For more information on the interaction between the Timeout and Failure Severity attribute and the virtual host failback policy, see "Virtual Hosts and Failover" on page 119.

To set the Timeout and Failure Severity attribute from the command line, use this option:

```
--probeSeverity nofailoverautorecover|noautorecover
```

Service Priority

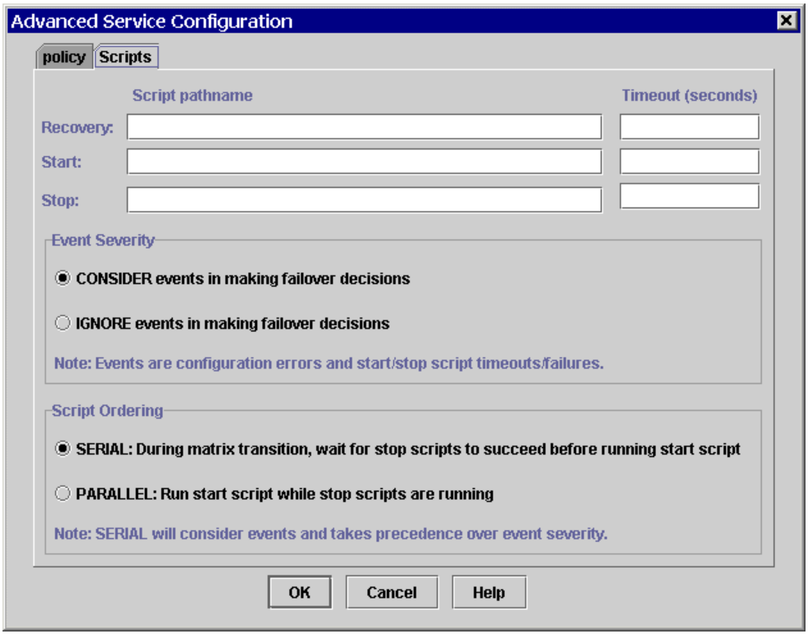
The service priority is used by HP Clustered File System when it fails over services. Service priorities are natural numbers, with 0 (zero) representing the highest priority and higher numbers representing lower priorities. If multiple failures prevent HP Clustered File System from placing a virtual host on a server where all of its associated services are available, HP Clustered File System next looks for a server where the associated service with the highest priority is available. By default, all service monitors have a priority of 0.

To set the Service Priority from the command line, use this option:

```
--priority <priority>
```

Custom Scripts

The Scripts tab lets you customize a service monitor with Start, Stop, and Recovery scripts. You can also configure the event severity and script ordering for the service monitor.


Advanced Service Configuration

	Script pathname	Timeout (seconds)
Recovery:		
Start:		
Stop:		

Event Severity

● CONSIDER events in making failover decisions

○ IGNORE events in making failover decisions

Note: Events are configuration errors and start/stop script timeouts/failures.

Script Ordering

● SERIAL: During matrix transition, wait for stop scripts to succeed before running start script

○ PARALLEL: Run start script while stop scripts are running

Note: SERIAL will consider events and takes precedence over event severity.

Scripts

Service monitors can optionally be configured with scripts that are run at various points during cluster operation. The script types are as follows:

Recovery script. Runs after a monitor probe failure is detected, in an attempt to restore the service.

Start script. Runs as a service is becoming active on a server.

Stop script. Runs as a service is becoming inactive on a server.

When a monitor is instantiated for a service (because the ClusterPulse daemon is starting or the configuration has changed), HP Clustered File System chooses the best server to make the service active. The Start script is run on this server. On all other servers configured for the monitor, the Stop script is run to ensure that the service is not active.

Start scripts must be robust enough to run when the service is already started, without considering this to be an error. Similarly, Stop scripts must be robust enough to run when the service is already stopped, without considering this to be an error. In both of these cases, the script should exit with a zero exit status.

This behavior is necessary because HP Clustered File System runs the Start and Stop scripts to establish the desired start/stop activity, even though the service may actually have been started by something other than HP Clustered File System before ClusterPulse was started. The Start and Stop scripts must also handle recovery from events that may cause them to run unsuccessfully.

For example, if the system runs out of swap space while running a Start script, the script will fail and exit non-zero. The service could then become active on another server, causing the Stop script to run on the original server even though the Start script did not complete successfully.

To configure scripts from the command line, use these options:

```
--recoveryScript <script>
--startScript <script>
--stopScript <script>
```

Event Severity

By default, HP Clustered File System treats the failure or timeout of a Start or Stop script as a failure of the associated monitored service and may initiate failover of the associated virtual hosts. Configuration errors can also cause this behavior.

Such a failure or timeout creates an event associated with the monitor on the server where the failure or timeout occurred. You can view these events on the Management Console and clear them from the Console or command line after you have fixed the problems that caused them.

You can configure the failover behavior with the Event Severity attribute. There are two settings:

CONSIDER. This is the default value. Events are considered when HP Clustered File System makes failover decisions.

IGNORE. Events are ignored and Start or Stop script failures will not cause failover. This is useful when the action performed by the Start and Stop scripts is not critical, but is important enough that you want to keep a record of it.

To configure event severity from the command line, use this option:

```
--scriptSeverity consider|ignore
```

Script Ordering

Script ordering determines the order in which HP Clustered File System runs Start and Stop scripts when a virtual host moves from one server to another. If you do not configure a monitor with Start and Stop scripts, the script ordering configuration has no effect. There are two settings:

SERIAL. This is the default setting. When a virtual host moves from one server to another, HP Clustered File System enforces the following strict ordering sequence for running Start and Stop scripts:

1. HP Clustered File System runs the Stop script on all servers where the virtual host should be inactive.
2. HP Clustered File System waits for all Stop scripts to complete.
3. HP Clustered File System runs the Start script on the server where the virtual host is becoming active.

PARALLEL. HP Clustered File System does not enforce the strict ordering sequence for Stop and Start scripts. The scripts run in parallel across the cluster as a virtual host is in transition.

The **PARALLEL** configuration can speed up failover time for services that do not depend on strict ordering of Start and Stop scripts. Assuming that it is safe to run the scripts in parallel (which depends on your application), this setting can also increase the chances of a successful failover because HP Clustered File System does not have to wait for the failing server to finish running its Stop script.

To configure script ordering from the command line, use this option:

```
--ordering serial|parallel
```

Other Configuration Procedures

On the management console, these procedures can be performed from the Servers, Virtual Hosts, or Applications tab. Locate the appropriate service monitor and then right-click to see the available procedures.

Delete a Service Monitor

Select the service monitor to be deleted, right-click, and select Delete.

To delete a service monitor from the command line, use this command:

```
mx service delete <vhost:port>
```

Disable a Service Monitor on a Specific Server

Select the service monitor to be disabled, right-click, and select Disable.

To disable a service monitor from the command line, use this command:

```
mx service disable <vhost:port> <server> ...
```

Enable a Previously Disabled Service Monitor

Select the service monitor to be enabled, right-click, and select Enable.

To enable a service monitor from the command line, use this command:

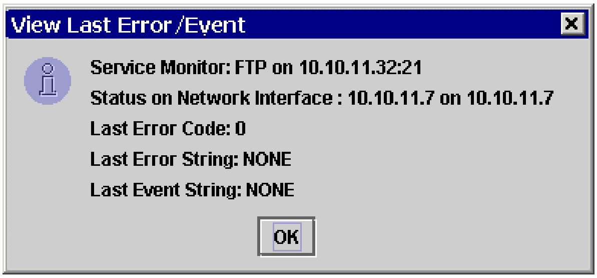
```
mx service enable <vhost:port> <server> ...
```

Remove Service Monitor from a Server

To remove a service monitor from a the network interface associated with a specific server, right-click, and select Remove From Server. You will then be asked to verify that you want to remove the server.

View Service Monitor Errors

To view the last error for a service monitor, select that service monitor, right-click, and select View Last Error.


View Last Error /Event

Service Monitor: FTP on 10.10.11.32:21

Status on Network Interface : 10.10.11.7 on 10.10.11.7

Last Error Code: 0

Last Error String: NONE

Last Event String: NONE

Clear Service Monitor Errors

Select the service monitor where the event occurred, right-click, and select Clear Last Event. To clear a monitor event from the command line, use this command:

```
mx service clear <vhost:port> <server> ...
```

11

Configure Device Monitors

HP StorageWorks Clustered File System provides built-in device monitors that can be used to watch disk devices or to monitor the status of PSFS filesystems. You can also create custom device monitors.

Overview

HP Clustered File System provides the following types of device monitors. To configure a device monitor, you will need to specify the probe timeout and frequency and a monitor-specific value.

Type	Probe Timeout	Probe Frequency	Other Values
SHARED_FILESYSTEM	Default: 5 seconds	Default: 30 seconds	Filesystem, filename
DISK	Default: 5 seconds	Default: 30 seconds	Partition to monitor
GATEWAY	5 seconds	5 seconds	Gateway IP address
CUSTOM	Default: 60 seconds	Default: 60 seconds	User probe script

Multi-Active Device Monitors

GATEWAY, DISK, and SHARED_FILESYSTEM monitors are multi-active. This means that they are active on multiple servers at the same time. This configuration allows the device monitor to probe a shared resource such as a network gateway or a PSFS filesystem. The status returned by the monitor probe determines the placement of the associated virtual hosts. For example, if a probe fails on the primary server for a virtual host, the virtual host may fail over to a backup server. See "Device Monitors and Failover" on page 142 for details about where a device monitor is active.

Custom device monitors can be configured to be either multi-active or single-active. With a single-active configuration, the monitor is active only on the primary node.

SHARED_FILESYSTEM Device Monitor

If you have created a virtual host for applications that obtain their data from a PSFS filesystem, or if you just want to monitor access to a PSFS filesystem, you can use the SHARED_FILESYSTEM device monitor to ensure that the filesystem is healthy and the virtual host remains available.

A SHARED_FILESYSTEM device monitor is associated with a specific PSFS filesystem. The monitor probes the filesystem at periodic intervals to determine the filesystem health. The probe attempts to open and read a file located on the filesystem. When you configure the monitor, you can specify the file to be read, or you can let HP Clustered File System create a file for this purpose.

A SHARED_FILESYSTEM device monitor attempts to be active on each server specified in its configuration. The device monitor probe reports an UP status for the filesystem only if it is mounted and file content can be read from the filesystem.

When you create a SHARED_FILESYSTEM device monitor, be sure to configure the following advanced options:

- Virtual hosts. Select any virtual hosts that should fail over if the monitor probe reports a DOWN status for the filesystem.

- Servers. Select all servers that have mounted the filesystem and are running the applications associated with the virtual hosts.

You might also want to create Start, Stop, or Recovery scripts to customize the behavior of the monitor.

GATEWAY Device Monitor

When certain network failures occur, the servers in a cluster can lose communication with each other. This situation can result in a partition, or split, of the cluster. For example, in a two-server cluster, each server would assume that it remained in the cluster and that the other server was down. The gateway device monitor detects the network failure and prevents the cluster from partitioning. The GATEWAY device monitor is multi-active and can be configured on multiple servers.

When you configure the monitor, you will be asked for the network address of a router or other gateway probe device. The IP address of the device should be on a different subnet than the servers in the cluster. The device should respond to a ping and should be visible on each server.

After the gateway device monitor is configured on a server, it pings the gateway device periodically. If a network failure occurs and the ping fails, any active virtual hosts on the server will become inactive and fail over to another server.

Disk and Custom Device Monitors

Disk Device Monitor

The DISK device monitor can be used to monitor a disk device. It periodically attempts to read the first block of the disk partition that you specify to determine whether the disk is operating normally. This device monitor is multi-active and can be configured on multiple servers.

Custom Device Monitor

A CUSTOM device monitor can be used when the built-in DISK type is not sufficient. Custom device monitors can be particularly useful when integrating HP Clustered File System with a custom application. Custom monitors can be configured to be either single-active or multi-active.

HP Clustered File System treats custom monitors just as it does the built-in DISK monitor, except that you must supply the probe script. In the script, probe commands should determine the health of the device as necessary. If the device is operating normally, the probe script should exit with exit status zero. If the device is not operating normally, the probe script should exit with a non-zero exit status. HP Clustered File System will interpret the non-zero exit status as a failure of the device, and take the appropriate action.

See "Advanced Topics" on page 166 for information about developing probe scripts for custom monitors and integrating monitors with custom applications.

Server Assignments

Typically a DISK or custom device monitor is configured for a single server, meaning that a private device is being monitored and the device monitor configuration is unique to a particular server. You may want to use a device monitor if you require an asymmetric monitor configuration across the cluster.

Failover

If a monitored device fails, HP Clustered File System attempts to relocate any virtual hosts that depend on the monitored device to a healthier server. A virtual host is never active on a server with an inactive, down, or disabled monitored device. If every server configured for a virtual host has a down device that the virtual host depends on, the virtual host will not be active anywhere in the cluster and thus will be totally down.

Device Monitors and Failover

When you create a device monitor, you specify a list of servers on which the device monitor will operate. The servers are placed in order: primary, backup #1, backup #2, and so on. ClusterPulse considers the status and enablement of the servers when determining where to activate a device monitor.

Device Monitor Activeness Policy

ClusterPulse uses the following device monitor activeness policy to determine the server or servers where it will make a device monitor active. The policy described here is accurate for this release but it may change in future releases.

The device monitor activeness policy decision is made as follows:

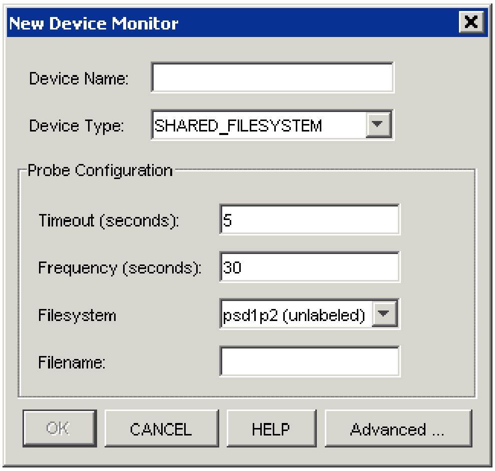
1. If the device monitor on a specific server is disabled, then the device monitor will not be made active on that server.
2. ClusterPulse considers the list of servers that are both up and enabled and that are configured for the device monitor. Note the following:
   - A server that has not finished joining the cluster (see "Server Access to the SAN" on page 179) is not considered up for the purpose of activating the device monitor.
   - A server is considered down if it loses coordinated communication with the cluster (for example, the server crashed or was shut down, HP Clustered File System was shut down on the server, the server failed to schedule a cluster group communication process for an extended period of time).
3. If the device monitor is multi-active, it will be active on all servers passing evaluation for steps 1 and 2. If the device monitor is single-active, then the list of servers is filtered by the following policy items to determine the one server where the device monitor will be active.
4. From the list of servers that pass evaluation for steps 1 and 2, ClusterPulse tries to find a server with all services up and enabled on the virtual hosts associated with this device monitor. If Clusterpulse finds a server meeting these conditions, it will use it, preferring servers earlier in the list of servers for this device monitor.
5. If there are no servers with completely healthy services, ClusterPulse picks a server that has at least one service up and enabled. If ClusterPulse finds a server meeting these conditions, it will use it, preferring services earlier in the list of servers for this device monitor.

6. If ClusterPulse cannot locate a server meeting these conditions, it will make the device monitor active on the first server in the list of servers for this device monitor.

Add or Modify a Device Monitor

Select the appropriate option:

- To add a new device monitor, select the server to be associated with the monitor from the Servers window, right-click, and select Add Device Monitor (or click the Device Monitor icon on the toolbar). Then configure the device monitor on the New Device Monitor window.
- To update an existing device monitor, select the monitor on the Servers window, right-click, and select Properties. Then make your changes on the Device Monitor Properties window.



New Device Monitor

Device Name:

Device Type: SHARED_FILESYSTEM

Probe Configuration

Timeout (seconds): 5

Frequency (seconds): 30

Filesystem psd1p2 (unlabeled)

Filename:

Device Name: Type the name of the device monitor. You can use up to 32 alphanumeric characters.

Device Type: Select the appropriate device type (DISK, GATEWAY, CUSTOM, or SHARED_FILESYSTEM). See “Overview” on page 139 for a description of these monitors.

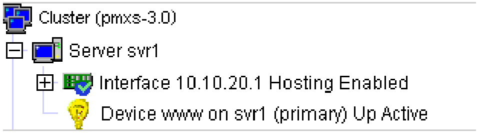
Frequency and **Timeout:** These fields are set to the default values for the type of device you have selected. Change them as needed.

The other monitor parameters are dependent on the type of monitor that you are creating.

- DISK monitor. At the "Monitor partition" prompt, specify a partition on the disk. The monitor will periodically attempt to read the first block of this partition to determine whether the disk is operating normally. Type the full partition name such as */dev/sda1*.
- CUSTOM monitor. At the "User probe script" prompt, specify the pathname to the probe script to be used with the monitor. See "Custom Scripts" on page 166 for information about writing probe scripts.
- SHARED_FILESYSTEM monitor. At the "Filesystem" prompt, select the filesystem that is to be monitored. The "Filename" parameter is optional. Enter the name of a file that the monitor probe should open and attempt to read to determine the health of the filesystem. The filename should be relative to the mount point of the filesystem. When the filesystem is mounted, the mountpath will be prepended to the filename to determine the complete filename path that should be probed.

  If you do not enter a filename, HP Clustered File System will create a file that will be used by the monitor probe. The file will be named *MxS/<serverIP>/<monitor_name>*, where *<serverIP>* is the dotted decimal IP address of the hostname for the server, and *<monitor_name>* is the name assigned to the SHARED_FILESYSTEM device monitor.
- GATEWAY monitor. At the "Gateway address" prompt, specify the IP address of the gateway device (such as a router). The IP address must be on a different subnet than the servers in the cluster.

The following example shows a SHARED_FILESYSTEM monitor created on the server svr1.


Cluster (pmxs-3.0)
Server svr1
Interface 10.10.20.1 Hosting Enabled
Device www on svr1 (primary) Up Active

To add or update a device monitor from the command line, use this command:

```
mx device add|update [--type CUSTOM|DISK|SHARED_FILESYSTEM]
[--timeout <seconds>] [--frequency <seconds>] [--parameters
<parameters>] [<other_arguments>] <devicename> ...
```

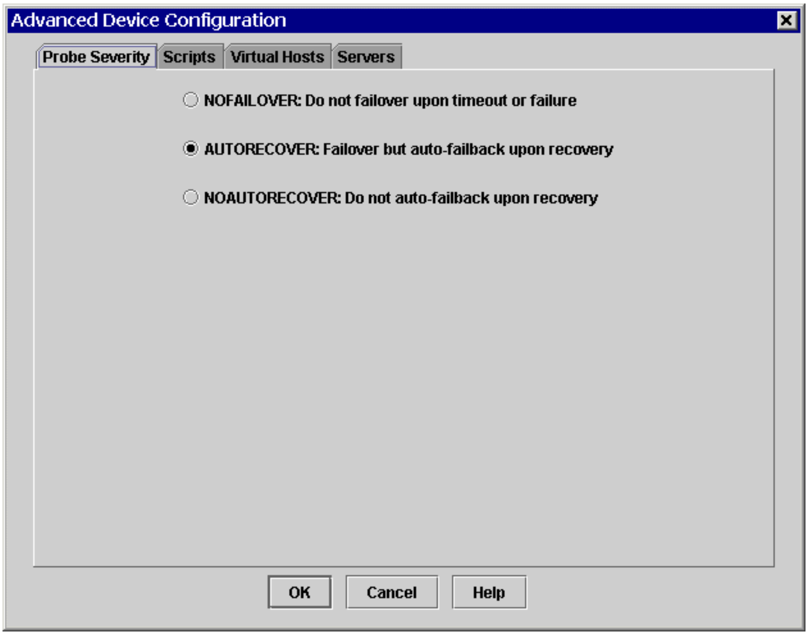
See “Advanced Settings for Device Monitors” for information about the other arguments that can be used for device monitors.

Advanced Settings for Device Monitors

You can use the advanced settings to fine-tune device monitors. To access the advanced settings, select the Advanced button on the Add Device Monitor or Device Monitor Properties window.

Probe Severity

The Probe Severity tab lets you specify the failover behavior of the device monitor.



Advanced Device Configuration

○ NOFAILOVER: Do not failover upon timeout or failure

● AUTORECOVER: Failover but auto-failback upon recovery

○ NOAUTORECOVER: Do not auto-failback upon recovery

The Probe Severity setting works with the virtual host policy (either AUTOFAILBACK or NOFAILBACK) to determine what happens when a monitored device fails.

The default policies (AUTOFAILBACK for the virtual host and AUTORECOVERY for the device monitor) cause ClusterPulse to fail over the associated virtual hosts to a backup network interface on another server when the monitor probe fails. When the device is restored, ClusterPulse fails back the virtual hosts to the network interface on the original server.

You can use the Probe Severity attribute to change this behavior. There are three settings for Probe Severity: NOFAILOVER, AUTORECOVER, and NOAUTORECOVER.

NOFAILOVER. When the monitor probe fails, ClusterPulse does not fail over to a backup network interface. This option is useful when the monitored resource is not critical, but is important enough that you want to keep a record of its health.

AUTORECOVER. This is the default. The virtual host fails over when a monitor probe fails. When device access is recovered on the original node, failback occurs according to the virtual host's failback policy.

NOAUTORECOVER. The virtual host fails over when a monitor probe fails and the monitor is disabled on the original node, preventing automatic failback. When the monitor is reenabled, failback occurs according to the virtual host's failback policy.

The NOAUTORECOVER option is useful when integrating HP Clustered File System with a custom application where certain application-specific actions must be taken before failback can occur.

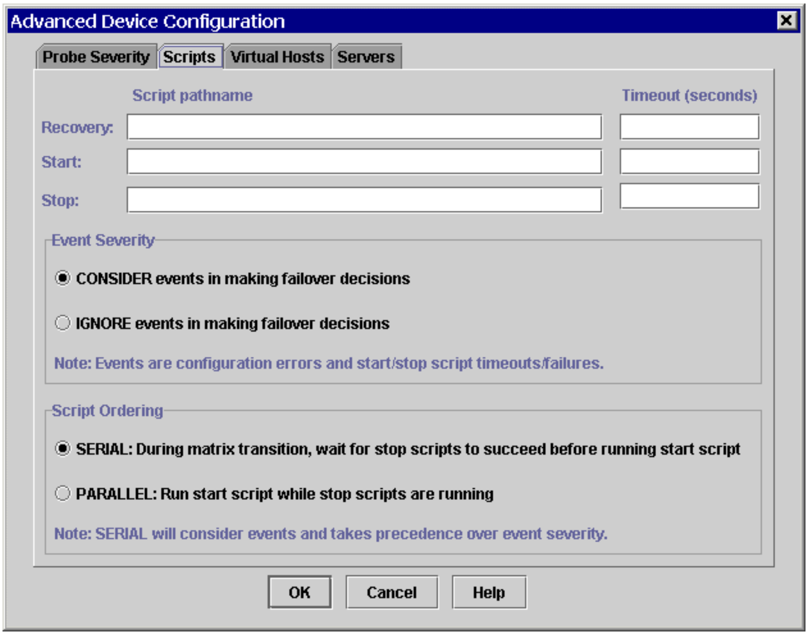
For more information on the interaction between the Probe Severity attribute and the virtual host failback policy, see "Virtual Hosts and Failover" on page 119.

To set the Probe Severity from the command line, use this option:

```
--probeSeverity nofailover|autorecover|noautorecover
```

Custom Scripts

The Scripts tab lets you configure custom Recovery, Start, and Stop scripts for a device monitor.



Advanced Device Configuration

	Script pathname	Timeout (seconds)
Recovery:		
Start:		
Stop:		

Event Severity

● CONSIDER events in making failover decisions

○ IGNORE events in making failover decisions

Note: Events are configuration errors and start/stop script timeouts/failures.

Script Ordering

● SERIAL: During matrix transition, wait for stop scripts to succeed before running start script

○ PARALLEL: Run start script while stop scripts are running

Note: SERIAL will consider events and takes precedence over event severity.

Device monitors can optionally be configured with scripts that are run at various points during cluster operation. The script types are as follows:

Recovery script. Runs after a monitor probe failure is detected, in an attempt to restore the device.

Start script. Runs as a device is becoming active on a server.

Stop script. Runs as a device is becoming inactive on a server.

When a monitor is instantiated for a device (because the ClusterPulse daemon is starting or the configuration has changed), HP Clustered File System chooses the best server to make the device active. The Start script is run on this server. The Stop script is run on all other servers configured for the monitor to ensure that the device is not active on those servers.

Start scripts must be robust enough to run when the device is already started, without considering this to be an error. Similarly, Stop scripts must be robust enough to run when the device is already stopped, without considering this to be an error. In both of these cases, the script should exit with a zero exit status.

This behavior is necessary because HP Clustered File System runs the Start and Stop scripts to establish the desired start/stop activity, even though the device may actually have been started by something other than HP Clustered File System before the ClusterPulse daemon was started. The Start and Stop scripts must also handle recovery from events that may cause them to run unsuccessfully.

For example, if the system runs out of swap space while running a Start script, the script will fail and exit non-zero. The device could then become active on another server, causing the Stop script to run on the original server even though the Start script did not complete successfully.

To configure scripts from the command line, use these options:

```
--recoveryScript <script>
--startScript <script>
--stopScript <script>
```

Event Severity

By default, HP Clustered File System treats the failure or timeout of a Start or Stop script as a failure of the associated monitored device and may initiate failover of the associated virtual hosts. Configuration errors can also cause this behavior.

Such a failure or timeout creates an event associated with the monitor on the server where the failure or timeout occurred. You can view these events in the Management Console and clear them using the Console or **mx device clear** command after you have fixed the problems that caused them.

You can configure the failover behavior with the Event Severity attribute. There are two settings:

CONSIDER. This is the default value. Events are considered when HP Clustered File System makes failover decisions.

IGNORE. Events are ignored and Start or Stop script failures will not cause failover. This is useful when the action performed by the Start and Stop scripts is not critical, but is important enough that you want to keep a record of it.

To configure event severity from the command line, use this option:

```
--scriptSeverity consider|ignore
```

Script Ordering

Script ordering determines the order in which HP Clustered File System runs Start and Stop scripts when a shared device or virtual host moves from one server to another. If you do not configure a monitor with Start and Stop scripts, the script ordering configuration has no effect.

There are two settings for script ordering.

SERIAL. This is the default setting. When a shared device or virtual host moves from one server to another, HP Clustered File System enforces the following strict ordering sequence for running Start and Stop scripts:

1. HP Clustered File System runs the Stop script on all servers where the shared device or virtual host should be inactive.
2. HP Clustered File System waits for all Stop scripts to complete.
3. HP Clustered File System runs the Start script on the server where the virtual host or shared device is becoming active.

PARALLEL. HP Clustered File System does not enforce the strict ordering sequence for Stop and Start scripts. The scripts run in parallel across the cluster as a shared device or virtual host is in transition.

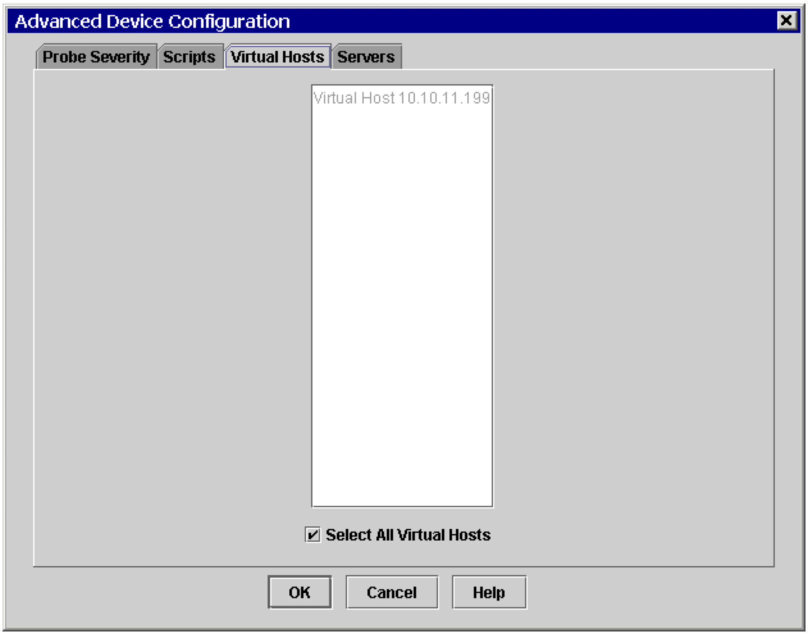
The **PARALLEL** configuration can speed up failover time for services and devices that do not depend on strict ordering of Start and Stop scripts. Assuming that it is safe to run the scripts in parallel (which depends on your application), this setting can also increase the chances of a successful failover because HP Clustered File System does not have to wait for the failing node to finish running its Stop script.

To configure script ordering from the command line, use this option:

```
--ordering serial|parallel
```

Virtual Hosts

The Virtual Hosts tab lets you specify the virtual hosts that will fail over if the device fails. The default is all virtual hosts configured on the server.



Virtual Host 10.10.11.199

☑ Select All Virtual Hosts

When a device monitor detects a failure, HP Clustered File System attempts to fail over the active virtual hosts associated with that device monitor. By default, all virtual hosts on the servers used with the device monitor are dependent on the device monitor. However, you can specify that only certain virtual hosts be dependent on the device monitor.

For example, you might have a DISK monitor for a disk containing Web and FTP files. If the disk fails, you want HP Clustered File System to fail over the virtual hosts for these services. The server might also provide mail service; however, the virtual host for this service is not dependent on the disk and should not fail over if the disk fails.

If you are creating a SHARED_FILESYSTEM monitor, select the virtual hosts associated with applications that access data on the PSFS filesystem being monitored.

To specify virtual hosts from the command line, use this option:

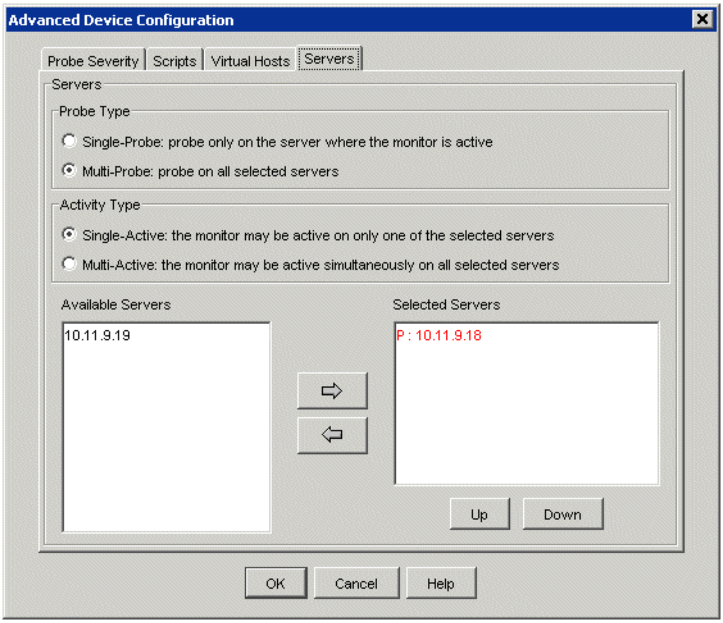
```
--vhosts <vhost1>,<vhost2>,...
```

Servers for Device Monitors

The Servers tab specifies the servers that use the device. The type of the device monitor determines how this window should be used.

- Typically a DISK device monitor is configured on one server.
- For SHARED_FILESYSTEM monitors, you should select the servers that mount the monitored filesystem and are running the applications that access data from that filesystem.
- Depending on their purpose, CUSTOM monitors can be configured on one or more servers. You can also select the Probe Type and Activity Type. The Probe Type specifies the servers on which the monitor probe will occur: either on only the server where the monitor is active, or on all servers configured for the monitor.

  The Activity Type specifies where the monitor can be active: either on one of the configured servers or on all of them. If you select Multi-Active, the probe type will be set to Multi-Probe and the Selected Servers column will not display P or B (for primary or backup).



Servers

Probe Type

○ Single-Probe: probe only on the server where the monitor is active

● Multi-Probe: probe on all selected servers

Activity Type

● Single-Active: the monitor may be active on only one of the selected servers

○ Multi-Active: the monitor may be active simultaneously on all selected servers

Available Servers: 10.11.9.19

Selected Servers: P : 10.11.9.18

To specify servers from the command line, use this option:

```
--servers <server1>,<server2>,...
```

To specify the Probe Type for a custom monitor, use this option:

```
--probe single|multiple
```

To specify the Activity Type, use this option:

```
--activity single|multiple
```

Other Configuration Procedures

On the management console, these procedures can be performed from the Servers or Applications tab. Locate the appropriate device monitor and then right-click to see the available procedures.

Delete a Device Monitor

Select the device monitor to be deleted from the Servers or Applications tab, right-click, and select Delete.

To delete a device monitor from the command line, use this command:

```
mx device delete <devicename> ...
```

Disable a Device Monitor

Select the device monitor to be disabled from the Servers or Application tab, right-click, and select Disable.

To disable a device monitor from the command line, use this command:

```
mx device disable <devicename> ...
```

Enable a Device Monitor

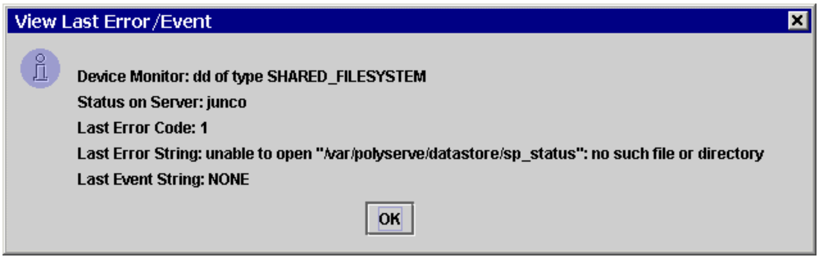
Select the device monitor to be enabled from the Servers or Application tab, right-click, and select Enable.

To enable a device monitor from the command line, use this command:

```
mx device enable <devicename> ...
```

View Device Monitor Errors

To view the last error that occurred on a device monitor, select that monitor on the Servers or Application tab, right-click, and select View Last Error.



View Last Error/Event

Device Monitor: dd of type SHARED_FILESYSTEM
Status on Server: junco
Last Error Code: 1
Last Error String: unable to open "/var/polyserve/datastore/sp_status": no such file or directory
Last Event String: NONE

Clear Device Monitor Errors

To clear a error from a device monitor, select that monitor on the Servers or Application tab, right-click, and select Clear Last Event.

To clear the device error from the command line, use this command:

```
mx device clear <devicename> <server> ...
```

12

Configure Notifiers

If you would like certain actions to take place when cluster events occur, you can configure notifiers that define how the events should be handled.

Overview

HP Clustered File System uses notifiers to enable you to view event information generated by servers, network interfaces, virtual hosts, service monitors, device monitors, and filesystems. Notifiers send events from these entities to user-defined notifier scripts. The events are categorized as State Transitions, Errors, Warnings, and Informational messages.

One server in a cluster holds the responsibility to forward a global event message to notifier scripts. The global events are generally state transition events. Some events such as errors, warnings, and informational messages from service monitors are reported only locally on a server. Any local event will be forwarded to notifier scripts that have subscribed to the event and entity combination on the server.

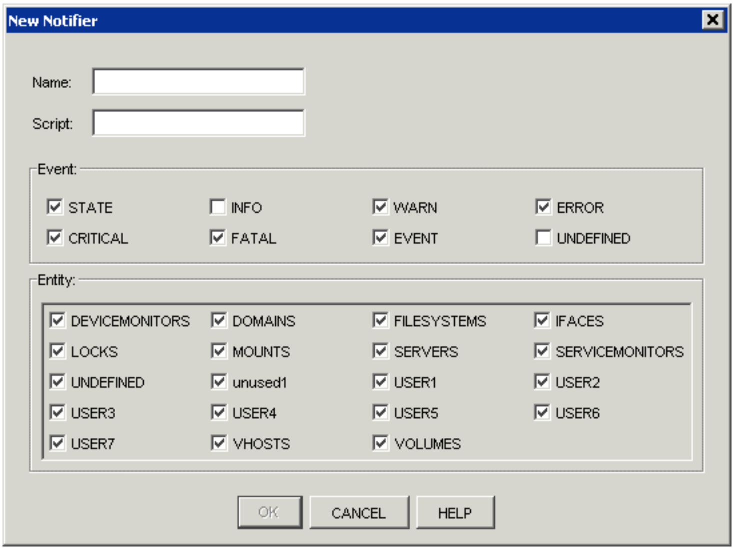
We recommend that you configure notifier scripts on each server in the cluster because the server responsible for handling global events may change in the normal course of ClusterPulse daemon processing. Only one notifier event message is delivered for each invocation of a notifier script. If multiple notifiers subscribe to the same combination of event and entity, then when an event is triggered, each notifier script is given a copy of the event message.

When adding a notifier, you will need to specify a name for the notifier and to supply the script to be run when an event is triggered that matches the event and entity combination. An event causes the script, which has its standard input wired to a pipe from the **notifier_agent**, to be run. The notifier script will be run with any arguments that you included in the script string. The script may read STDIN to accept the event message.

Add or Modify a Notifier

Select the Notifier tab on the Management Console to display the Notifiers window and then choose the appropriate option:

- To add a new notifier, select Cluster > Notifier > New (or click the Add Notifier icon on the toolbar) and then configure the notifier on the New Notifier window.
- To modify an existing notifier, select that notifier from the Notifiers window, right-click, and select Properties. Then make your changes on the Notifier Properties window.



Name:

Script:

Event:

☑ STATE	□ INFO	☑ WARN	☑ ERROR
☑ CRITICAL	☑ FATAL	☑ EVENT	□ UNDEFINED

Entity:

☑ DEVICEMONITORS	☑ DOMAINS	☑ FILESYSTEMS	☑ IFACES
☑ LOCKS	☑ MOUNTS	☑ SERVERS	☑ SERVICEMONITORS
☑ UNDEFINED	☑ unused1	☑ USER1	☑ USER2
☑ USER3	☑ USER4	☑ USER5	☑ USER6
☑ USER7	☑ VHOSTS	☑ VOLUMES	

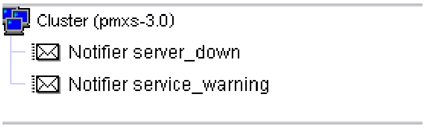
Name: Enter a name for the notifier. You can use up to 32 alphanumeric characters.

Script: Enter the name of the script that will be run when an event occurs.

Event: Check the events for which you want to receive notification.

Entity: Check the entities for which you want to receive notification. The USER1 - USER7 entities are user-defined entities for the **mxlogger** command. See "Add Your Own Messages to the Cluster Log File" on page 205.

The notifier now appears in the Notifiers window.



Cluster (pmxs-3.0)
Notifier server_down
Notifier service_warning

To add or modify a notifier from the command line, use this command:

```
mx notifier add|update --script <script> [--event {STATE|INFO|
WARN|ERROR},...] [--entity {SERVERS|IFACES|VHOSTS|
SERVICEMONITORS|DEVICEMONITORS|FILESYSTEMS}] <notifier>
```

The **--script** argument is optional when you are updating a notifier.

Other Configuration Procedures

Delete a Notifier

Select the notifier to be deleted from the Notifiers window, right-click, and select Delete.

To delete a notifier from the command line, use this command:

```
mx notifier delete <notifier> ...
```

Disable a Notifier

Select the notifier to be disabled from the Notifiers window, right-click, and select Disable.

To disable a notifier from the command line, use this command:

```
mx notifier disable <notifier> ...
```

Enable a Notifier

Select the notifier to be enabled from the Notifiers window, right-click, and select Enable.

To enable a notifier from the command line, use this command:

```
mx notifier enable <notifier> ...
```

Test a Notifier

Select the notifier to be tested from the Notifiers window, right-click, and select Test. The event messages for each configured entity will now be sent to the notifier.

To test a notifier from the command line, use this command:

```
mx notifier test <notifier> ...
```

Notifier Message Format

Notifier messages have the following format:

```
<Source> <Event> <Entity> <Event code number> <Date and
time> <Text>
```

Source: The source IP address of the server that generated the global event, or the string "LOCAL" for events generated locally on that server.

Event: One of **State**, **Error**, **Warn**, or **Info**.

Event code number: Either a three-digit number representing the event or the number 0 (zero). The number 0 means that the message was generated by a test of a notifier.

Date and Time: Represented as *mmm dd yyyy hh:mm:ss*.

Text: The full text description of the event. The text can include specific cluster information that can cause the format to vary significantly.

Sample Notifier Messages

Following is an example of a notifier message:

```
10.10.1.1 State VHOSTS 130 Oct 31 2000 13:13:00 Virtual
host change - 10.1.1.1 now active on 10.10.1.1
```

The Test Notifier option causes a test event to be generated for each of the event/entity combinations that you configure for the notifier. Following is an example:

```
10.10.1.1 Error SERVICEMONITORS 0 Oct 31 2000 13:31:31
TEST Notifier message for <notifier name>
```

A Sample Notifier Script

The following script resets a switch when a notifier event:entity rule is true. For example, it could be run when another script places WARN:USER1 in the cluster log file.

.

```
#!/bin/bash

SW_ETC=/etc/hpcfs

login=admin
password=password

for host in $(cat $SW_ETC/psSAN.cfg  | awk -F\| '$1 ~
/^LOCAL/{print $3}'); do
# Hot reset each switch
(echo $login; sleep 1; echo $password; echo admin cancel;
sleep 1; echo y; sleep 1; echo admin start; echo hotreset;
sleep 2; echo y; sleep 30 )| \
telnet $host
done

exit 0
```

13

Test Your Configuration

After you have configured HP Clustered File System, we recommend that you perform a set of basic tests to validate that SAN shared filesystem operation, virtual host operation and failover, DNS load-balancing operation and failover, and failover of the LAN administrative network work correctly.

After completing these tests successfully, you may want to run a more substantial test of your specific requirements to validate that HP Clustered File System is working in your environment. We strongly recommend that you perform the first two sets of tests ("Test SAN Shared Filesystem Operation" and "Test Virtual Host Operation and Failover") to validate the basic configuration of the cluster.

Test SAN Shared Filesystem Operation

The first step in validating the successful operation of HP Clustered File System is to verify that your SAN hardware configuration, operating system configuration, and HP Clustered File System software configuration correctly support PSFS filesystems shared across the servers in your cluster.

Test SAN Connectivity and Shared Filesystem Operation

Use the following procedure to test basic SAN connectivity and shared filesystem operation in your cluster:

1. From the Management Console, log into one of the cluster servers.
2. Import an unused SAN disk into your cluster configuration.
3. Create a PSFS filesystem on an unused partition on this disk.
4. Mount the PSFS filesystem on each server in the cluster. This filesystem will now be shared by these servers.
5. Log into each of the servers as *root*. From the shell prompt, perform some basic I/O tests to the shared filesystem using a tool such as **tar**. Verify that changes to the filesystem made by each server are visible to all other servers in the cluster.

Test Filesystem Failure Recovery

1. Configure the cluster with a shared filesystem as described in the previous procedure.
2. Pull the LAN network connection(s) from one of the servers. Verify that this server is unable to access the shared filesystem. Verify that the other servers in the cluster are still able to access the shared filesystem.
3. Reconnect the LAN network to the server and then reboot it. Verify that the server, upon rebooting, is able to mount the shared filesystem. Verify that all servers are able to access the shared filesystem.
4. Pull the SAN network connection(s) from one of the servers. Verify that this server is unable to access the shared filesystem. Verify that the other servers in the cluster are still able to access the shared filesystem.
5. Reconnect the SAN network to the server and then reboot it. Verify that the server, upon rebooting, is able to mount the shared filesystem. Verify that all servers are able to access the shared filesystem.

6. Power off one of the servers. Verify that the other servers in the cluster are still able to access the shared filesystem.
7. Restore the power to the server and then reboot it. Verify that this server, upon rebooting, is able to mount the shared filesystem. Verify that all servers are able to access the shared filesystem.

Test Virtual Host Operation and Failover

The following procedure tests automatic failover and recovery reintegration. It is best to run these tests in a non-production environment.

Test Failure and Reintegration of Servers

Use the following procedure to test server failure, failover, and reintegration:

1. From the Management Console, log into a backup cluster server.
2. Configure the cluster with a single virtual host.
3. Validate that all servers are up, that the virtual host is active on the primary server, and that the backup servers are inactive.
4. Pull the LAN network connection(s) from the primary server.
5. Verify that the Management Console shows that the primary server is down and inactive and that the first backup server is active.
6. Reconnect the primary server to the LAN.
7. Verify that the primary server re-enters the cluster as the active host and that the first backup server becomes inactive.

Test Failure and Reintegration of Service Monitors

The following procedure tests the failure and reintegration of a service monitor:

1. From the Management Console, log into a backup server.
2. Add a service monitor to the virtual host already defined in the system.

3. Verify that all servers are up, that the service you are testing is up, and that the virtual host is active on the primary server and inactive on the backup servers.
4. Stop the service you are testing on the primary server (for example, for HTTP, bring down the HTTP daemon).
5. Verify that HP Clustered File System detects the service failure. The virtual host should be inactive on the primary server and active on the first backup server.
6. Start the service that you are testing on the primary server.
7. Verify that HP Clustered File System detects that the service has become active.
8. Verify that the virtual host is active on the primary server and inactive on the backup servers.

We strongly recommend that you run these tests successfully before attempting more advanced tests. ISPs and hosting services should also attempt multiple virtual host failover and reintegration tests.

Test DNS Load-Balancing and Failover

Testing the correct functionality of the load-balancing configuration is a little more difficult than the primary/backup failover configuration. There are more components to test, including the DNS server and multiple clients.

The easiest way to test load-balancing and configuration is to use a single client from outside of the cluster, ping the virtual hostnames, and determine whether the DNS server is providing the virtual host IP addresses in succession. The DNS server must be set up as described in “Configure Servers for DNS Load Balancing” on page 41.

The following examples assume that you have a two-server round robin setup, but you can easily extend this test for more than two servers.

Assume that you have the following DNS round robin setup:

www.acmd.com	192.168.100.1
www.acmd.com	192.168.100.2

The DNS name is www.acmd.com and 192.168.100.1 is a virtual host with primary on acmd1 and backup on acmd2. 192.168.100.2 is primary on acmd2 and backup on acmd1. DNS is set up to round robin on the servers acmd1 and acmd2, using the virtual host addresses 192.168.100.1 and 192.168.100.2.

Validate Correct Load-Balancing Operation

The following procedure validates that DNS round robin and HP Clustered File System are working correctly.

1. Ping www.acmd.com.
2. Validate that the returned IP address is 192.168.100.1.
3. Ping www.acmd.com.
4. Validate that the returned IP address is 192.168.100.2.
5. Continue to ping and validate that the opposite IP address is returned by DNS.

Validate Load-Balancing When a Server Is Down

This test verifies that the load-balanced cluster operates correctly when one of the servers is down or removed from the cluster. The latter part of the test validates that the server begins to handle its portion of the load-balanced traffic when it rejoins the cluster.

1. From the HP Clustered File System console, log into one of the servers in the cluster.
2. Pull the LAN network connection(s) from a server that you are not logged onto.
3. Verify that HP Clustered File System detects that the server is down and that the backup for this virtual host takes over operation from the primary.
4. Ping the address www.acmd.com multiple times.
5. Verify that all requests are still being returned correctly.
6. Restore the LAN network connections.

7. Verify that DNS now serves up both IP addresses again and that the ping is returned correctly by both.

Test LAN Failover of Administrative Cluster Traffic

Use the following procedure to test the LAN administrative traffic failover capability of HP Clustered File System:

1. Connect your cluster servers with at least two physically separate LANs. Configure the Linux network software to enable the interfaces to these networks on each of the cluster servers.
2. From the Management Console, log into one of the cluster servers.
3. Observe on the Management Console which of your LANs is currently being used for cluster administrative traffic. Physically disconnect one of the server's LAN connections to this network.
4. Observe on the Management Console that HP Clustered File System has noticed the network disconnection and has failed over the administrative cluster traffic to one of the other LANs.
5. Reconnect the disconnected server to the LAN. Observe on the Management Console that HP Clustered File System has noticed that network connectivity has been restored for this LAN.

14

Advanced Topics

The topics described here provide technical details about HP Clustered File System operations. This information is not required to use HP Clustered File System in typical configurations; however, it may be useful if you want to design custom scripts and monitors, to integrate HP Clustered File System with custom applications, or to diagnose complex configuration problems.

Custom Scripts

When you create a custom service or device monitor, you will need to supply a probe script that can determine whether the service or device is up or down.

You can also supply Recovery, Start, and Stop scripts for any service or device monitor. Start and Stop scripts are run during monitor virtual host transitions. The Recovery script is run only when a probe script fails.

Requirements for Custom Scripts

The following requirements apply to all custom scripts created for use with service and device monitors.

- The specification of the command and its parameters must be less than 512 bytes in total length.
- The script must be an executable file in the filesystem. You can use shell scripts, Java scripts, C programs, or other types of executable files.

- The command will be executed as *root*.
- For a service monitor, the file must be installed on each server associated with the virtual host on which the service monitor is located.
- For a device monitor, the file must be installed on each server that is configured with the virtual hosts associated with the device monitor.
- The command termination exit status is used to signal script success or failure. A 0 (zero) exit status indicates script success; any other exit status indicates script failure.

Types of Custom Scripts

You can develop the following types of custom scripts for use with service and device monitors.

Probe Scripts

Probe scripts run periodically and check the health of a service or device. You will need to determine what the probe should look at and how it should exit.

For example, you might want to create a custom service monitor for an application that creates a pid file when it is running. You could write a shell script called **/usr/local/myprobes/testpid** that looks up the pid from the application’s pid file and then determines whether the pid is actually running.

Here is the **testpid** script:

```
if [ ! -e $1 ]; then   # Does the pid file exist?
       exit 1
fi
read pid < $1        # Read the pid into the variable pid
kill -0 $pid         # Does the pid exist in the process table?
```

When you create the service monitor, you would specify the following as the user probe script. (Specifying the application’s pid file as a parameter instead of naming it in the script allows the **testpid** script to be used with any application that creates a pid file.)

```
usr/local/myprobes/testpid /var/run/application.pid
```

When the monitor executes the **testpid** script, it will first determine whether the */var/run/application/pid* file exists. If the file does not exist, the script exits with a non-zero exit status, which the monitor interprets as a failure.

If the file does exist, the script reads the pid from the file into the variable pid. The **kill** command then determines whether the pid is running. The exit status of the **kill** command is the exit status of the script.

If the kill command finds that the pid is running, it will exit with status 0, and the script will exit with status 0. The monitor will interpret the 0 exit status as "success" and will signal to the cluster that the application is up.

If the kill command finds that the pid is not running, it will exit with a non-zero status, and the script will exit with that same status. The monitor will interpret that exit status as "failure," which will signal the monitor that the application is down. HP Clustered File System will then take the action configured for the service monitor, which is typically to fail over the virtual host associated with the monitor.

When you create the custom service or device monitor for the probe script, you can set both the frequency at which the probe script should be executed and the timeout period, which is the maximum amount of time that the **monitor_agent** daemon will wait for the probe to complete.

You can create more elaborate probe scripts as necessary. The key points are to check whether the service or device is up and then to return a corresponding exit status. The service or device monitor uses only the exit status to determine whether the probe succeeded or failed, with 0 indicating success and any other value indicating failure.

Recovery Scripts

A Recovery script runs after a monitor probe fails. The script attempts to restore the service and prevent failover of the virtual host(s) associated with the monitor.

Recovery scripts are useful if there is an automatic way to recover from a common failure mode for an application.

For example, if you are monitoring an application called **myservice** that is normally started at boot time, but which is buggy and crashes occasionally, you could use a Recovery script to reduce the frequency of failovers. The script could contain the following line:

```
/etc/rc.d/init.d/myservice restart
```

When you add a recovery script to a service or device monitor, you can set a timeout period, which is the maximum amount of time that the **monitor_agent** daemon will wait for the Recovery script to complete.

Start and Stop Scripts

These scripts are run when a monitor is instantiated for a service (because either the ClusterPulse daemon is starting or the configuration has changed). The scripts establish the desired start/stop activity. The Start script is run on the server where the monitor will be active. The Stop script is run on all other servers.

Following are some typical uses of the Start and Stop scripts:

- An application requires the ownership of a shared resource to be effective. The Start script tries to take ownership of the resource (returning non-zero if it fails to do so), and the Stop script yields ownership. Be sure that script ordering is strict, which is the default. (Script ordering is an advanced configuration option and is set on the Scripts tab.)
- Ensure the availability of non-shared resources. For example, a Start script can start an auxiliary process needed by the monitored application if it is not already running.
- Perform cleanup tasks such as killing any unreaped children of a failed application process. Stop scripts can be used for this purpose.

In some cases the monitored service or device is actually started by something other than HP Clustered File System before ClusterPulse is started. The Start script must be robust enough to run in this circumstance without considering it to be an error. Similarly, Stop scripts must be robust enough to run when the service is already stopped, without considering this to be an error.

Start and Stop scripts must also handle recovery from events that could cause them to run unsuccessfully. For example, the system might run out of swap space while running a Start script, causing the script to fail and exit non-zero. The service could then become active on another server, causing the Stop script to run on the original server even though the Start script had not completed successfully.

When you add Start and Stop scripts to a service or device monitor, you can set a timeout period for each script.

Script Environment Variables

When you specify a script for a service or device monitor, HP Clustered File System sets the following environment variables for that script.

MX_METHOD=(START|STOP|RECOVER|PROBE)

The type of script (Start, Stop, Recovery, or probe).

MX_ACTIVE_STATE=(ACTIVE|INACTIVE)

Whether the script is being run on an active instance of the object. For example, if a service-monitor script is being run on the server that currently has the active virtual host, its state will be ACTIVE. For device-monitor scripts, the state is ACTIVE if the monitored device is active on the server running the script.

MX_SERVER=IP address

The primary address of the server that calls the script. The address is specified in dotted decimal format.

MX_TYPE=(SERVICE|DEVICE)

Whether the script is for a service or device monitor.

MX_VHOST=IP address

The IP address of the virtual host. The address is specified in dotted decimal format. (Applies only to service monitors.)

MX_PORT=Port or name

The port or name of the service monitor. (Applies only to service monitors.)

MX_NAME=name

The name of the device monitor. (Applies only to device monitors.)

HP Clustered File System does not set any other variables. If a script requires a variable such as a pathname, it will need to set it.

The Effect of Monitors on Virtual Host Failover

Typically a virtual host has a primary network interface and one or more backup network interfaces. On the servers supplying the interfaces, the state of the virtual host is either active or inactive.

When HP Clustered File System needs to fail over a virtual host, it looks for the healthiest backup server. To make this determination, HP Clustered File System considers the state of any service or device monitors associated with the virtual host.

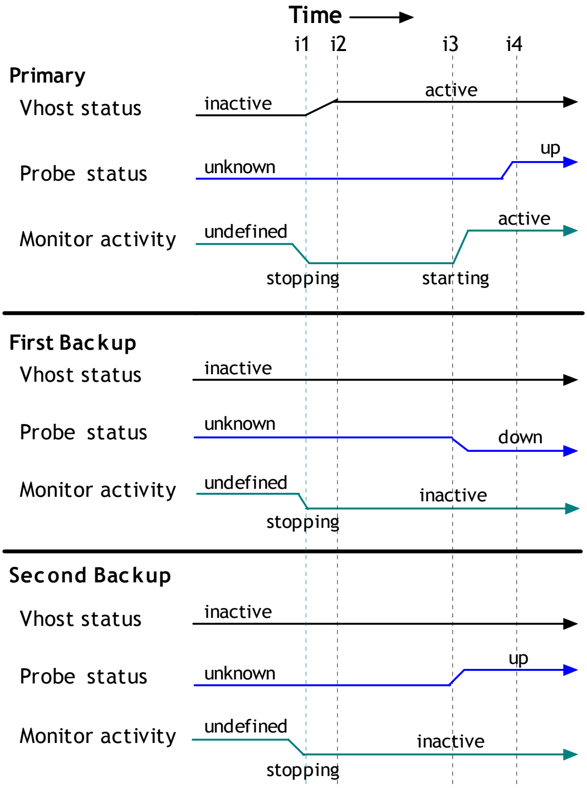
Service Monitors

Service monitors are configured on a virtual host. The service to be monitored is expected to be running simultaneously on the primary server and on the backup servers associated with the virtual host. The monitor uses a probe mechanism to determine whether the service is up. The probe mechanism is in one of the following states on each server: Up, Down, Unknown, Timeout.

A service monitor also has an activity status on each server. The status can be one of the following: Starting, Active, Suspended, Stopping, Inactive, Failure.

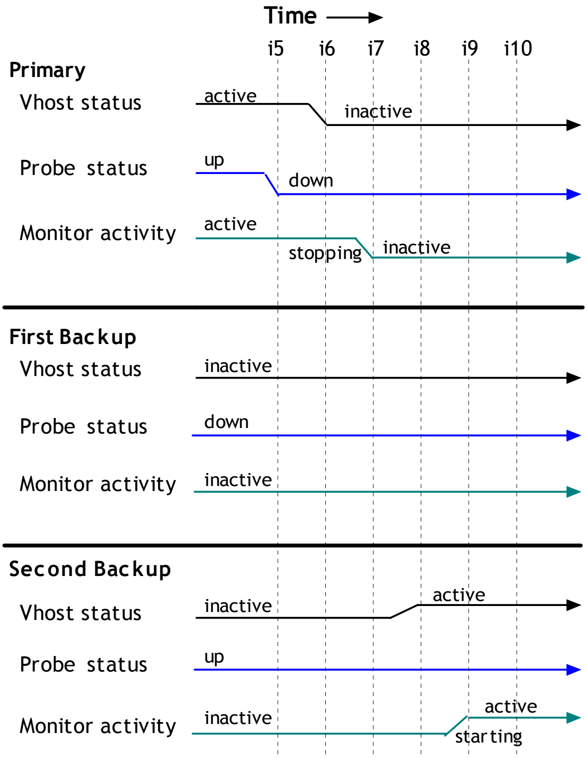
The following examples show state transitions for a service monitor that uses the default values for autorecovery, priority, and serial script ordering. Start and Stop scripts are also defined for the monitor. The virtual host associated with the monitor has a primary interface and two backup interfaces.

The first example shows the state transitions that occur at startup from an unknown state. At i1, all instances of the monitor have completed stopping. At i2, the virtual host is configured on the Primary. At i3, the monitor start script begins on the Primary and probing begins on the backups. At i4, probing begins on the Primary.



When a failure occurs on the Primary, the virtual host needs to fail over to a backup. HP Clustered File System now looks for the best location for the virtual host. Because the probe status on the first backup is “down,” HP Clustered File System chooses the second backup, where the probe status is “up.”

At i5 in the following example, the probe fails on the Primary. At i6, the virtual host is deconfigured on the Primary. At i7, the monitor stop script begins on the Primary. At i8, the virtual host is configured on the second backup. At i9, the monitor start script begins on the second backup. At i10, probing begins on the second backup.



Custom Device Monitors

A custom device monitor is associated with a list of servers and a list of virtual hosts configured on those servers. A custom device monitor can be active on only one server at a time. On each server, the monitor uses a probe mechanism to determine whether the service is active. The probe mechanism is in one of the following states on each server: Up, Down, Unknown, Timeout.

A custom device monitor also has an activity status on each server. This status indicates the current activity of the monitor on the server. The status can be one of the following: Starting, Active, Suspended, Stopping, Inactive, Failure.

If it is necessary to fail over a virtual host associated with the device monitor, HP Clustered File System looks for a server that meets both of these conditions: the device monitor is active, and the device monitor probe reports an Up status.

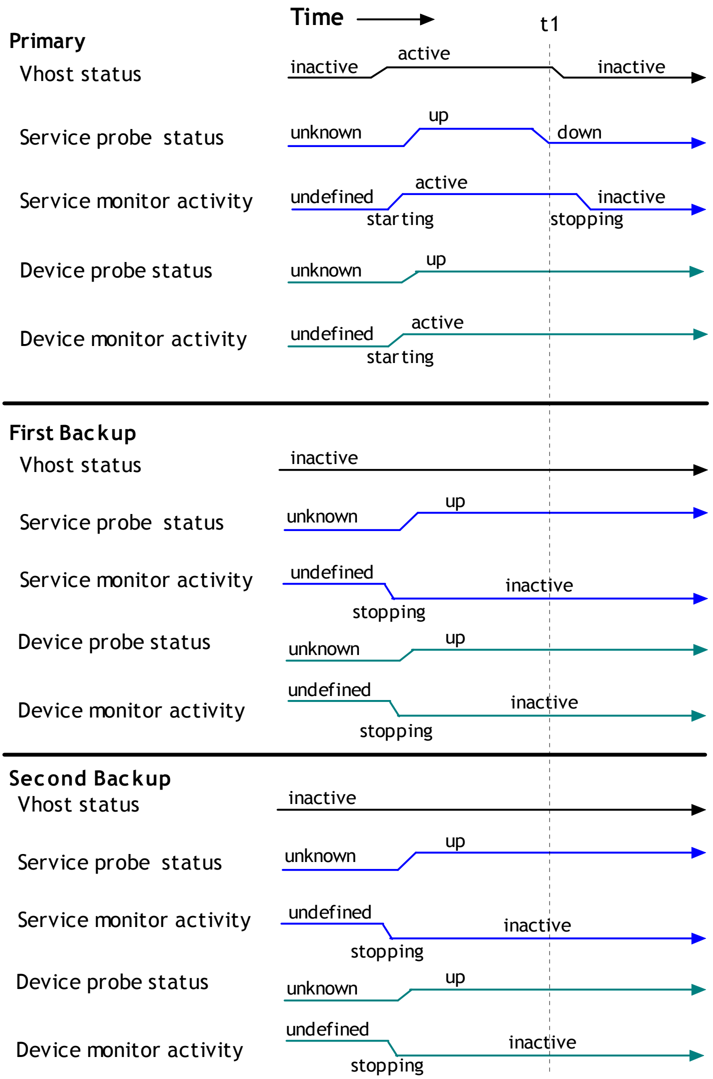
If HP Clustered File System cannot locate a server where the device is active and the probe status is Up, the virtual host will be left inactive across the cluster.

The following example shows the state transitions for a custom device monitor and its associated virtual host. The device monitor uses the default values for autorecover, priority, and serial script ordering, and has a Start and Stop script defined. The virtual host is configured with a Primary network interface and two backup interfaces. There is also a service monitor defined on the virtual host. It uses the default values for autorecovery, priority, and serial script ordering and has Start and Stop scripts defined.

The example begins at initial startup from an unknown state and the virtual host is then located on the primary interface.

At time t1, the service monitor probe on the Primary reports a Down status. The virtual host then becomes inactive all servers.

The virtual host cannot be active on the Primary server because the custom service monitor is Down. It cannot become active on either of the backup servers because the custom device monitor associated with the virtual host is Inactive on these servers.


Time
t1
Primary
Vhost status
inactive
active
inactive
Service probe status
unknown
up
down
Service monitor activity
undefined
active
starting
inactive
stopping
Device probe status
unknown
up
Device monitor activity
undefined
active
starting
First Backup
Vhost status
inactive
Service probe status
unknown
up
Service monitor activity
undefined
inactive
stopping
Device probe status
unknown
up
Device monitor activity
undefined
inactive
stopping
Second Backup
Vhost status
inactive
Service probe status
unknown
up
Service monitor activity
undefined
inactive
stopping
Device probe status
unknown
up
Device monitor activity
undefined
inactive
stopping

Integrate Custom Applications

There are many ways to integrate custom applications with HP Clustered File System:

- Use service monitors or device monitors to monitor the application
- Use a predefined monitor or your own user-defined monitor
- Use Start, Stop, and Recovery scripts

Following are some examples of these strategies.

Device Monitor or Service Monitor?

When deciding whether to monitor your custom application with a service monitor or a device monitor, first determine whether the application provides a service over the network directly to the outside world. If so, you should consider using a service monitor. Because a service monitor is associated with a virtual host, it represents an IP address provided to the outside world to access your services.

For example, one common use of HP Clustered File System is to increase the availability of Web servers. The availability of the HTTP service is best modeled by a service monitor associated with the virtual host to which outside users connect. HP Clustered File System provides a built-in HTTP service monitor for this use. The other built-in service monitors offer more examples of services which are best modeled by service monitors.

Alternatively, if you want to monitor something associated with a physical server, you should consider using a device monitor. Monitoring the health of a local disk drive, for example, is the job of a device monitor. The built-in DISK device monitor is available for this purpose.

Similarly, if you have a Web server that communicates with a back-end database, you might want to devise your own user-defined device monitor to determine whether each server in the cluster has a good network connection to the database server.

Built-In Monitor or User-Defined Monitor?

To decide whether to use a built-in monitor or a user-defined monitor, first determine whether a built-in monitor is available for the service you want to monitor and then consider the degree of content verification that you need. If you are monitoring a custom application for which a built-in monitor is not available but the application provides network services through certain ports, you may consider it sufficient to use built-in TCP monitors on each of those ports.

However, if you want to verify that the application is still accepting connections to the port and is also producing reasonable output, you should use a user-defined service monitor. You will need to write a probe script that connects to the port and then tests how the script responds to various commands.

NOTE: The user-defined monitor dialog prompts you for a service monitor name and not a port because you may be writing a monitor for an application that does not provide network services and therefore needs no port.

A Sample Custom Monitor

This example uses service monitors with a custom application called **myservice**. This application provides some facilities to clients who connect to port 2468 and speak a protocol. You have already set up a virtual host called *vh1* for the IP address to which external clients are going to connect. How do you make a service monitor for this application?

The simplest way is to use a generic built-in TCP monitor on port 2468. This monitor verifies that it is possible to connect to port 2468, which probably indicates most of the time that the application is functioning. However, a problem might occur that causes the application to continue accepting connections but not produce meaningful output. To detect this situation, you will need a more complex and robust monitor involving a script written with a utility such as **expect**(1).

This script connects to port 2468, sends a string specified by the protocol, and determines whether it has received an expected response.

You distribute this script to the same location on all servers on virtual host *vh1* and then create a CUSTOM service monitor, specifying the path of the script as the “user probe script” parameter. This provides not only verification of the connection, but a degree of content verification.

The CUSTOM monitor can also include Start and Stop scripts. Suppose the **myservice** application caches transactions induced by requests from external users for later commitment to a back-end database server. You want to ensure that if a failover occurs, any transactions previously acknowledged by one server appear as complete to users connecting to the new server. This is the kind of situation where Start and Stop scripts are useful.

The Stop script should commit all cached transactions to the back-end database. The Start script should force an update of your view of the database from the back-end server.

Then, if the service moves from the primary server to a backup, all transactions performed on the primary will be sent to the database and from there to the backup server before connections are directed to the backup server.

Limit the Servers That Can Join a Cluster

The network secret key controls server access to the cluster. When you create a cluster configuration, you will specify a password that is used as the network secret key. Servers must have this password to join the cluster. When you add a new server to the cluster and import the cluster configuration to it, the network secret key will be included. See the *HP Clustered File System Setup Guide* for more information.

15

SAN Maintenance

The following information and procedures apply to SANs used with HP StorageWorks Clustered File System.

Server Access to the SAN

When a server is either added to the cluster or rebooted, HP Clustered File System needs to take some administrative actions to make the server a full member of the cluster with access to the shared filesystems on the SAN. During this time, the Management Console reports the message "Joining cluster" for the server. The message is removed when the administrative actions are complete and the server is participating in the cluster.

A server can lose access to all or part of the SAN for several reasons:

- Network problems. Cluster partitioning can occur when cluster network communications are lost but the affected server can still access the SAN. HP Clustered File System then removes the server's access to the SAN to protect the integrity of the shared filesystems. You will need to reboot the server to restore its SAN access.
- SAN hardware problems or misconfiguration of a FibreChannel switch. The server's access to the SAN will be restored automatically when the hardware is repaired or reconfigured correctly.
- Repeated I/O errors when the server tries to write to a PSFS journal. The server then loses access to the affected filesystem.

When the disk experiencing the I/O errors is fixed, the server will automatically regain access to the filesystem.

The Management Console typically displays an alert message when a server loses access to the SAN. (See Appendix B for more information about these messages.)

Membership Partitions

HP Clustered File System uses a set of membership partitions to control access to the SAN and to store the device naming database, which includes the global device names for SAN disks imported into the cluster. Typically, the membership partitions are created when you install HP Clustered File System.

When a membership partition is placed under HP Clustered File System control, write access is disabled to prevent accidental corruption from other applications.

The following commands can be useful in diagnosing or repairing problems with the membership partitions:

- The **mxsanlk** command reports the status of the SAN ownership locks used to control access to the SAN. These locks are stored on the membership partitions.
- The **mxmpconf** command starts a user interface that can be used to create a new set of membership partitions or to repair the existing partitions.

Display the Status of SAN Ownership Locks

HP Clustered File System uses a set of disk-based data structures called SANlocks to protect filesystem integrity. If a problem causes a cluster to split into two or more network partitions, the SANlocks ensure that only one of the resulting network partitions has access to the SAN.

Each SANlock is stored in a membership partition. Before a cluster can begin accessing the SAN, it must first acquire a majority of the SANlocks. The SANlocks are acquired in order.

mxsanlk displays the status of the SANlock stored in each membership partition. It can be used to determine whether any of the membership partitions need to be repaired. Also, if a network partition occurs, **mxsanlk** can be used to determine which network partition has control of the SAN.

Following is some sample output. The command was issued on host 10.10.30.3. The SDMP administrator is the administrator for the cluster to which the host belongs. There are three membership partitions.

```
# mxsanlk
This host:                      10.10.30.3
This host’s SDMP administrator: 10.10.30.1

Membership Partition    SANlock State
--------------------    -------------
/dev/rpsd/psd1p1         held by SDMP administrator
/dev/rpsd/psd2p1         held by SDMP administrator
/dev/rpsd/psd3p3         held by SDMP administrator
```

Any of these messages can appear in the “SANlock State” column.

- held by SDMP administrator

  The SANlock was most recently held by the SDMP administrator of the cluster to which the host where **mxsanlk** was run belongs.

- trying to lock, last held by host *X.X.X.X*

  The SANlock was most recently held by host *X.X.X.X* and may still be held by that host. The host on which **mxsanlk** was run is trying to acquire the SANlock.

- cannot access

  The host on which **mxsanlk** was run is unable to access the SANlock.The membership partition may need to be repaired.

- trying to lock, cannot access

  The host on which **mxsanlk** was run is trying to acquire the SANlock but is unable to access it. The membership partition may need to be repaired.

- locked, cannot access

  The host on which **mxsanlk** was run held the SANlock but is now unable to access it. The membership partition may need repair.

- trying to lock, not yet committed by owner

  The SANlock is either not held or has not yet been committed by its holder. The host on which **mxsanlk** was run is trying to acquire the SANlock.

- unlocked, trying to lock

  The SANlock does not appear to be held. The host on which **mxsanlk** was run is trying to acquire the SANlock.

- unlocked

  The SANlock does not appear to be held. If a host holds the SANlock, it has not yet committed its hold.

- initiating sdmp, not yet examined

  This is a transitional state. It indicates that the sdmp process responsible for the SANlock has been started but has not yet accessed the SANlock.

- sdmp process hung

  The SDMP process responsible for the SANlock is unresponsive.

- trying to lock, sdmp process hung

  The host on which **mxsanlk** was run is trying to acquire the SANlock but the SDMP process responsible for the SANlock is unresponsive.

- locked, sdmp process hung

  The host on which **mxsanlk** was run held the SANlock but the SDMP process responsible for the SANlock is now unresponsive.

- lock is corrupt, will repair

  This transitional state occurs after the SDMP has detected that the SANlock has been corrupted but before it has repaired the SANlock.

- trying to lock (lock is corrupt, will repair)

  The host on which **mxsanlk** was run is trying to acquire the SANlock. The SANlock was corrupted but will be repaired.

- locked (lock is corrupt, will repair)

  The host on which **mxsanlk** was run holds the lock. The SANlock was corrupted but will be repaired.

If a membership partition cannot be accessed, use the **mxmpconf** program to correct the problem.

When you invoke **mxsanlk**, it checks for the Storage Device Monitor Pulse (SDMP) daemon. This daemon is responsible for grabbing and maintaining the locks on the membership partitions. Depending on the status of the SDMP daemon, you may see one of the following messages:

```
Checking for SDMP activity, please wait...
Still trying...
The SDSMP is inactive at this host.
The SDMP appears to be inactive at this host.
```

If the SDMP daemon is not responding on the host, wait a few seconds and retry the command. If the command continues to fail, shut down the cluster and then restart it. This step should restart the SDMP daemon. If you continue to have problems, contact HP Technical Support.

Manage Membership Partitions with mxmpconf

Each server in the cluster has a membership partition file, which is called the “local MP list.” This file specifies the locations of the membership partitions. Each membership partition also has its own MP list. Because the membership partitions control access to the SAN, it is important that each server in the cluster have the same view of where the partitions are located.

However, it is possible for these views to get out of sync if changes to the membership partition configuration are made when all servers are not accessible. Therefore, before a server can negotiate for control of the SAN, it must have a local MP list that matches the membership lists stored on each of the membership partitions, and each membership partition must be in the Active state.

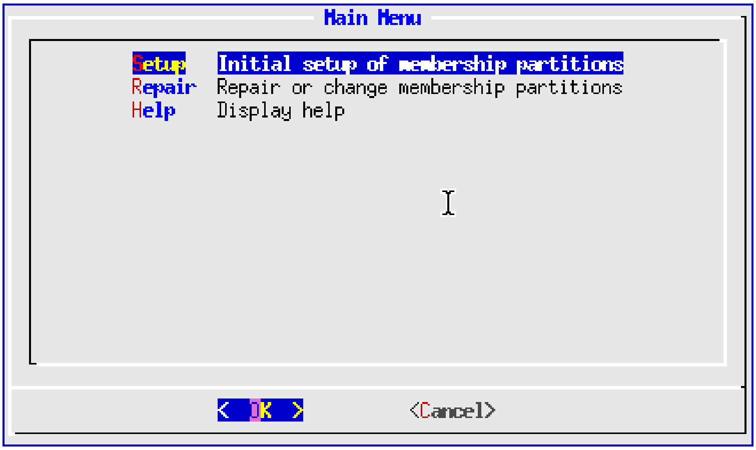
The **mxmpconf** utility can be used to repair any problems if a failure causes servers to have inconsistent views of the membership partitions.

The **mxmpconf** utility starts an ASCII interface that you can use to create a new set of membership partitions or to repair the existing partitions.

NOTE: HP Clustered File System cannot be running when you use **mxmpconf**. To stop the cluster, use the following command:

```
# /etc/init.d/pmxs stop
```

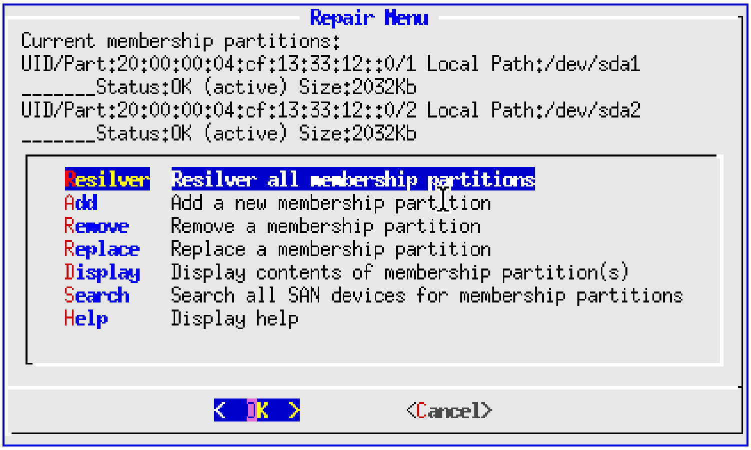
After stopping HP Clustered File System, type **mxmpconf** at the operating system prompt. The Main Menu is then displayed.


Main Menu
Setup Initial setup of membership partitions
Repair Repair or change membership partitions
Help Display help
< OK >
<Cancel>

Use the arrow keys to move between the options on the menu, or type the first letter of an option to move to that option. Use the Tab key to move between OK and Cancel. Press the Enter key to activate an option. Press the spacebar to go to the next page of an output display.

Maintain Membership Partitions with the Repair Option

The Repair Menu allows you to view the membership partition configuration and to perform several maintenance activities.



```
                            Repair Menu
 Current membership partitions:
 UID/Part:20:00:00:04:cf:13:33:12::0/1 Local Path:/dev/sda1
 _______Status:OK (active) Size:2032Kb
 UID/Part:20:00:00:04:cf:13:33:12::0/2 Local Path:/dev/sda2
 _______Status:OK (active) Size:2032Kb

     Resilver  Resilver all membership partitions
     Add       Add a new membership partition
     Remove    Remove a membership partition
     Replace   Replace a membership partition
     Display   Display contents of membership partition(s)
     Search    Search all SAN devices for membership partitions
     Help      Display help

                  <  OK  >          <Cancel>
```

The Repair Menu lists the current membership partitions according to the membership file maintained on the server where you are running the utility. Each server in the cluster has a membership partition file, which is called the "local MP list." Each SAN disk containing a membership partition also has its own list of the membership partitions. Under normal operations, these lists should all match.

Membership Partition Status

The status reported for each membership partition will be one of the following:

OK. The membership partition is included in the local membership partition list. This is the normal status.

NOT FOUND. The **mxmpconf** utility cannot find the device containing the membership partition.

INACCESSIBLE. The **mxmpconf** utility cannot access the device containing the membership partition.

CORRUPT. The partition is not valid.

MISMATCH. The membership partition is valid but its MP list does not match the server's local MP list.

If the status is NOT FOUND or INACCESSIBLE, there may be a problem with the disk or with another SAN component. When the problem is repaired, the status should return to OK.

If the status is CORRUPT, you should resilver the partition. This step copies the membership data from a valid membership partition to the corrupted partition.

NOTE: The membership partition may have become corrupt because it was used by another application. Before resilvering, verify that it is okay to overwrite any existing data on the partition.

If the status is MISMATCH, you will need to determine which membership partitions are correct (the partitions specified in the local MP list, or the partition labeled MISMATCH) and resilver accordingly.

Active and Inactive Membership Partitions

A membership partition can be either active or inactive. The current membership partitions should all be active. If there are old membership partitions in the cluster, you may want to either remove them or mark them as inactive.

A cluster could have an old membership partition if the disk containing it becomes inaccessible and you then add another partition in its place. When the disk access problem is corrected, the old membership partition will still be on the disk and will be marked active.

When a membership partition is active, the disk containing the partition cannot be imported into the cluster. You may see this situation if you want to import a disk that was previously used by another cluster. To fix this problem, use the Inactivate option on the Repair menu to change the state of the membership partition to "inactive." You can then import the disk into the cluster.

Sizes for Membership Partitions

HP Clustered File System stores the size of the smallest membership partition that was created during the HP Clustered File System configuration. When you add or replace a membership partition, the new partition must be at least as large as that original partition. For example, if you originally created 2-GB, 3-GB, and 4-GB membership partitions, any membership partitions created later on must be at least 2 GB in size.

Export Configuration Changes

When you change the membership partition configuration with **mxmpconf**, it updates the membership list on the local server. It also updates the lists on the disks containing the membership partitions specified in the local MP file.

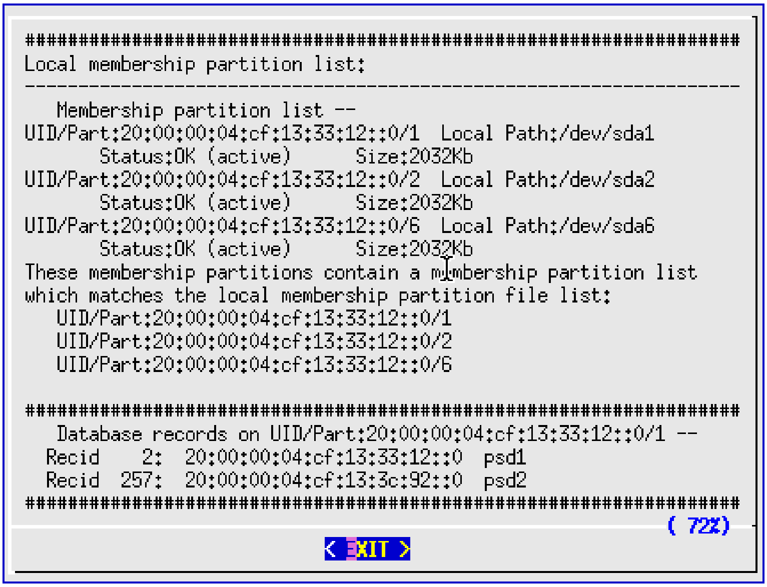
When you exit **mxmpconf**, you will be asked whether you want to export the new membership configuration to the other servers in the cluster. This step updates the local MP files on those servers.

Repair Menu Options

HP Clustered File System can operate with either one membership partition or three membership partitions. To ensure that a membership partition is always available, we recommend that you use three membership partitions.

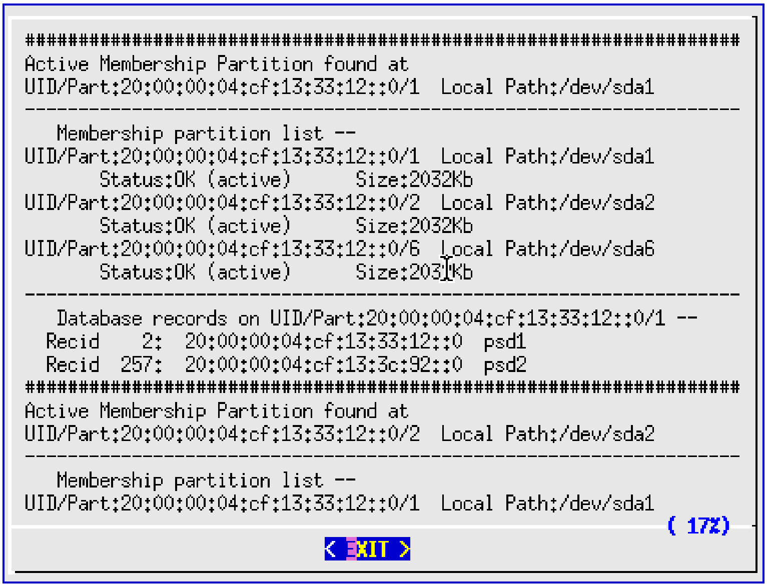
NOTE: When you exit the Repair menu, **mxmpconf** checks the number of membership partitions that are in effect after your changes. If you have added or removed membership partitions and there are now only two membership partitions, you will be returned to the Repair menu so that you can either add the third membership partition or remove one of the existing partitions.

Display Membership Partitions. The Display option lists information about membership partitions and can also copy the data to a file. The option shows the local membership partition list on the server where you are running **mxmpconf**. It then compares this list with the lists located on the disks containing the membership partitions. In the following example, the lists match. The output also includes the device database records for the disks containing the membership partitions.



```
###################################################################
Local membership partition list:
-------------------------------------------------------------------
   Membership partition list --
UID/Part:20:00:00:04:cf:13:33:12::0/1  Local Path:/dev/sda1
       Status:OK (active)      Size:2032Kb
UID/Part:20:00:00:04:cf:13:33:12::0/2  Local Path:/dev/sda2
       Status:OK (active)      Size:2032Kb
UID/Part:20:00:00:04:cf:13:33:12::0/6  Local Path:/dev/sda6
       Status:OK (active)      Size:2032Kb
These membership partitions contain a membership partition list
which matches the local membership partition file list:
   UID/Part:20:00:00:04:cf:13:33:12::0/1
   UID/Part:20:00:00:04:cf:13:33:12::0/2
   UID/Part:20:00:00:04:cf:13:33:12::0/6

###################################################################
   Database records on UID/Part:20:00:00:04:cf:13:33:12::0/1 --
  Recid    2:  20:00:00:04:cf:13:33:12::0  psd1
  Recid  257:  20:00:00:04:cf:13:3c:92::0  psd2
###################################################################
```

Search the SAN for Membership Partitions. The Search option searches the SAN for all partitions that appear to be membership partitions. You can also copy this data to a file.



```
###################################################################
Active Membership Partition found at
UID/Part:20:00:00:04:cf:13:33:12::0/1  Local Path:/dev/sda1
-------------------------------------------------------------------
   Membership partition list --
UID/Part:20:00:00:04:cf:13:33:12::0/1  Local Path:/dev/sda1
       Status:OK (active)      Size:2032Kb
UID/Part:20:00:00:04:cf:13:33:12::0/2  Local Path:/dev/sda2
       Status:OK (active)      Size:2032Kb
UID/Part:20:00:00:04:cf:13:33:12::0/6  Local Path:/dev/sda6
       Status:OK (active)      Size:2032Kb
-------------------------------------------------------------------
   Database records on UID/Part:20:00:00:04:cf:13:33:12::0/1 --
  Recid    2:  20:00:00:04:cf:13:33:12::0  psd1
  Recid  257:  20:00:00:04:cf:13:3c:92::0  psd2
###################################################################
Active Membership Partition found at
UID/Part:20:00:00:04:cf:13:33:12::0/2  Local Path:/dev/sda2
-------------------------------------------------------------------
   Membership partition list --
UID/Part:20:00:00:04:cf:13:33:12::0/1  Local Path:/dev/sda1
```

The output includes each membership partition found by the search, whether the partition is active or inactive, the membership list on the disk containing the partition, and the database records for the partitions.

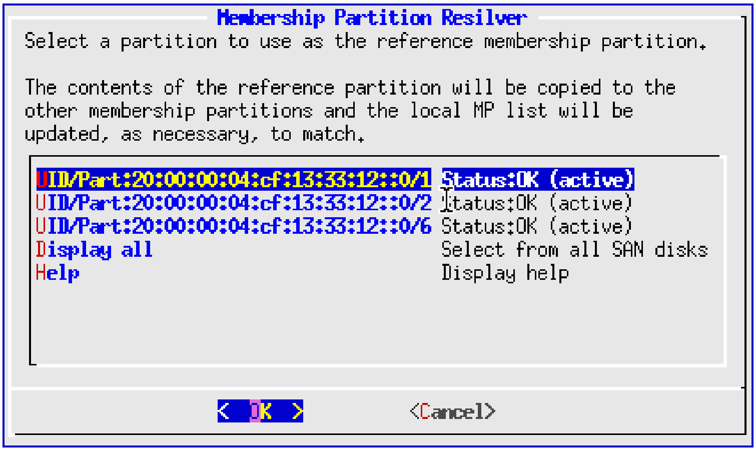
Resilver Membership Partitions. Typically, HP Clustered File System writes data to one membership partition and then copies, or resilvers, that data to the other membership partitions. Resilvering takes place automatically when a change is made to the device naming database or when you add, remove, or replace a membership partition.

However, in certain situations you may need to perform the resilver operation manually. For example, a membership partition might become corrupt or a local membership list might become out of date.

When you select the Resilver option, you will be asked to select the partition from which data should be copied.

When you choose a partition, the local path to that partition will appear at the bottom of the window.

NOTE: If you resilver from a partition that is labeled MISMATCH, the operation may initialize partitions that are not currently membership partitions; any existing data on those partitions will be overwritten. Use the Display option to see the membership partition lists for the current membership partitions.

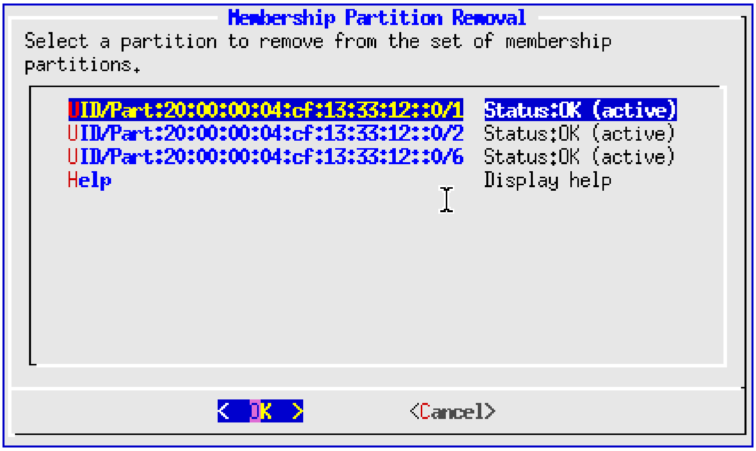


```
Membership Partition Resilver
Select a partition to use as the reference membership partition.

The contents of the reference partition will be copied to the
other membership partitions and the local MP list will be
updated, as necessary, to match.

UID/Part:20:00:00:04:cf:13:33:12::0/1 Status:OK (active)
UID/Part:20:00:00:04:cf:13:33:12::0/2 Status:OK (active)
UID/Part:20:00:00:04:cf:13:33:12::0/6 Status:OK (active)
Display all                           Select from all SAN disks
Help                                  Display help

              <  OK  >            <Cancel>
```

To resilver from a partition that is not in the local MP list, select Display all, which shows all disks in the SAN. When you select a disk, the partitions on that disk are displayed. To select a partition, move to that partition and press the spacebar. You can use the Search option on the Repair menu to locate a valid membership partition.

The resilver operation synchronizes all other membership partitions and the local membership partition list.

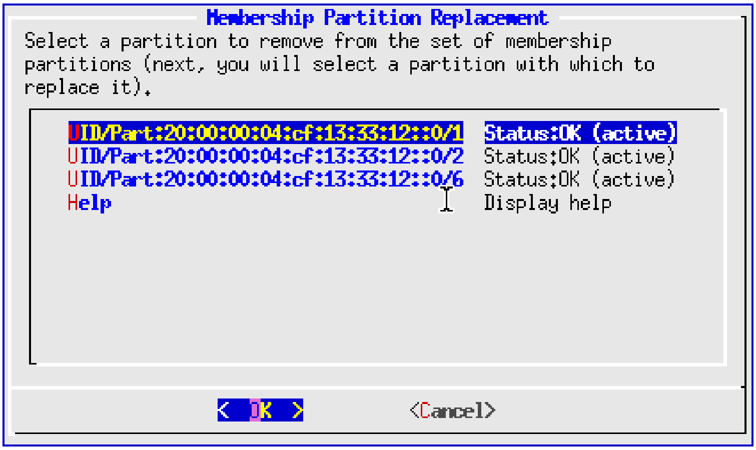
Remove a Membership Partition. The Remove option allows you to remove an existing membership partition.



```
Membership Partition Removal
Select a partition to remove from the set of membership
partitions.

    UID/Part:20:00:00:04:cf:13:33:12::0/1  Status:OK (active)
    UID/Part:20:00:00:04:cf:13:33:12::0/2  Status:OK (active)
    UID/Part:20:00:00:04:cf:13:33:12::0/6  Status:OK (active)
    Help                                   Display help
```

The remove operation removes the membership data and formatting from the partition. It also deletes the partition from the local MP list and from the membership partition lists on the remaining membership partitions.

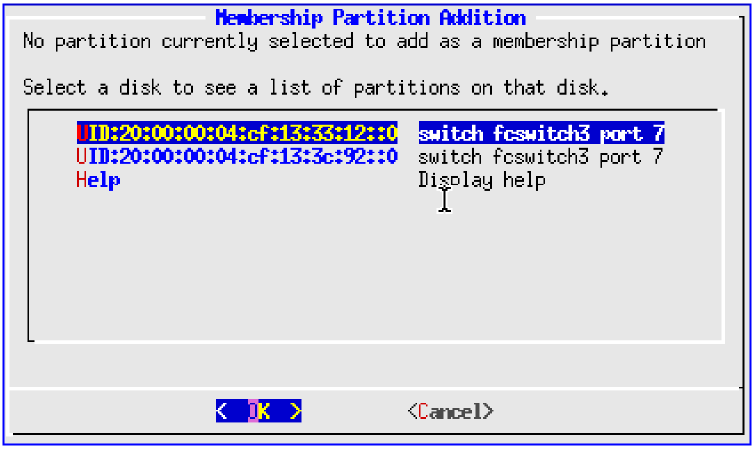
Replace a Membership Partition. The Replace option allows you to remove an existing membership partition and then select another partition to replace it. All membership partitions must currently have a status of OK.



```
Membership Partition Replacement
Select a partition to remove from the set of membership
partitions (next, you will select a partition with which to
replace it).

    UID/Part:20:00:00:04:cf:13:33:12::0/1  Status:OK (active)
    UID/Part:20:00:00:04:cf:13:33:12::0/2  Status:OK (active)
    UID/Part:20:00:00:04:cf:13:33:12::0/6  Status:OK (active)
    Help                                   Display help
```

After you select the partition to be removed, you will be asked to select the SAN disk containing the replacement partition. The partitions on that disk are then displayed. To select a partition, move to that partition and press the spacebar. When you choose the new partition, the local path to that partition will appear at the bottom of the window. Select Done to complete the operation.

Add a Membership Partition. If there are fewer than three membership partitions, the Repair Menu will include an option to add a new membership partition. To use this option, all membership partitions must have a status of OK.



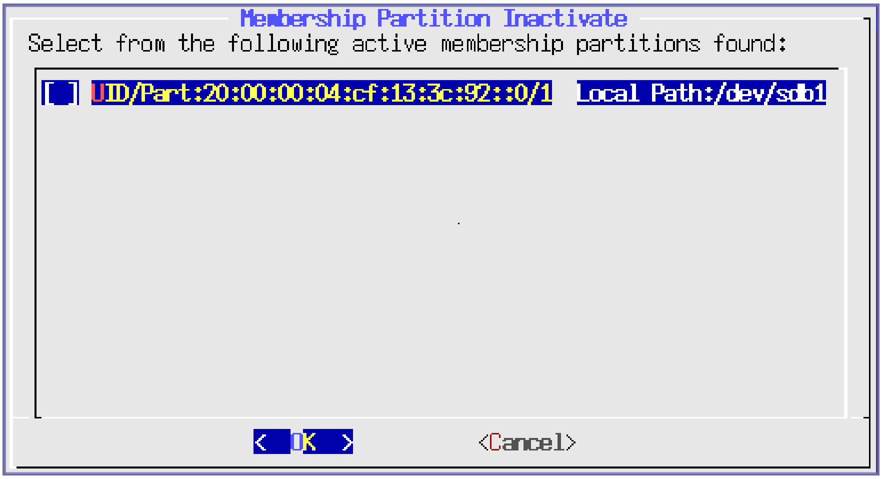
```
                Membership Partition Addition
 No partition currently selected to add as a membership partition

 Select a disk to see a list of partitions on that disk.

       UID:20:00:00:04:cf:13:33:12::0   switch fcswitch3 port 7
       UID:20:00:00:04:cf:13:3c:92::0   switch fcswitch3 port 7
       Help                             Display help
```

The Add option asks you to select the SAN disk containing the new partition. The partitions on that disk are then displayed. To select a partition, move to that partition and press the spacebar. (A 1-MB partition is adequate.) The local path to that partition then appears at the bottom of the window. Select Done to complete the operation.

The new partition is configured as a membership partition, the data on the existing membership partitions is copied to it, and the local membership partition list and the lists on the existing membership partitions are updated.

Inactivate a Membership Partition. This option can be used to inactivate membership partitions that are marked as active but are not part of the current set defined by the membership partition list.

When you select the Inactivate option, the Membership Partition Inactivate window lists all of the active membership partitions that **mxmpconf** has located. Select the partitions be to inactivated from this window.



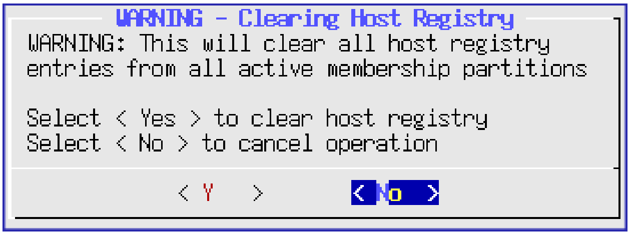
```
Membership Partition Inactivate
Select from the following active membership partitions found:
[ ] UID/Part:20:00:00:04:cf:13:3c:92::0/1  Local Path:/dev/sdb1
```

This option is useful if the cluster includes old membership partitions that are marked active or if you want to import a disk that contains an active membership partition.

Clear the Host Registry. This option removes all entries from the server registry. It should be used only in situations such as the scenario described under “Administration When Cluster Is in a Degraded State” on page 197.

CAUTION: Before clearing the server registry, be sure to reboot or power off any servers that were previously removed from the cluster and no longer had access to the SAN. After the servers have been rebooted, they can safely access the SAN. (If the servers are not rebooted, it is possible for them to corrupt filesystems.)

When you select the option to clear the host registry, the following warning will be displayed. Select Y to clear the registry, or No to cancel the operation.



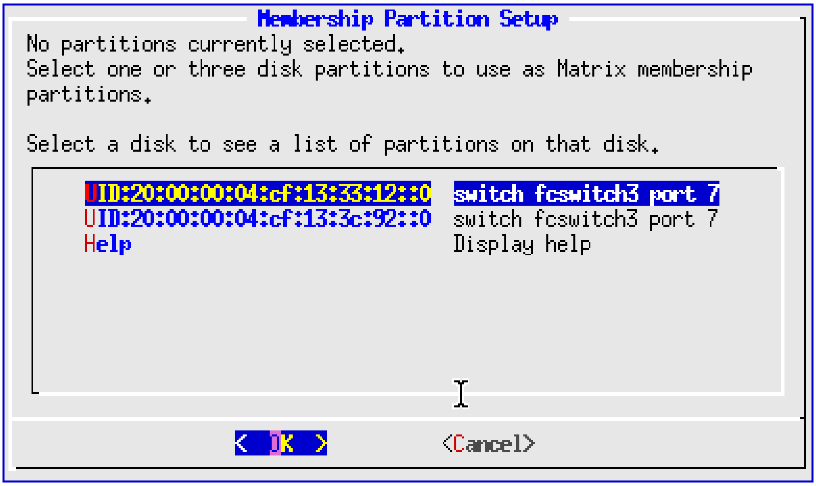
```
WARNING - Clearing Host Registry
WARNING: This will clear all host registry
entries from all active membership partitions

Select < Yes > to clear host registry
Select < No > to cancel operation

        < Y   >        < No  >
```

Create New Membership Partitions with the Setup Option

The Setup option allows you to create a new set of membership partitions. Typically, you will have done this when you installed HP Clustered File System and should not need to use this option.

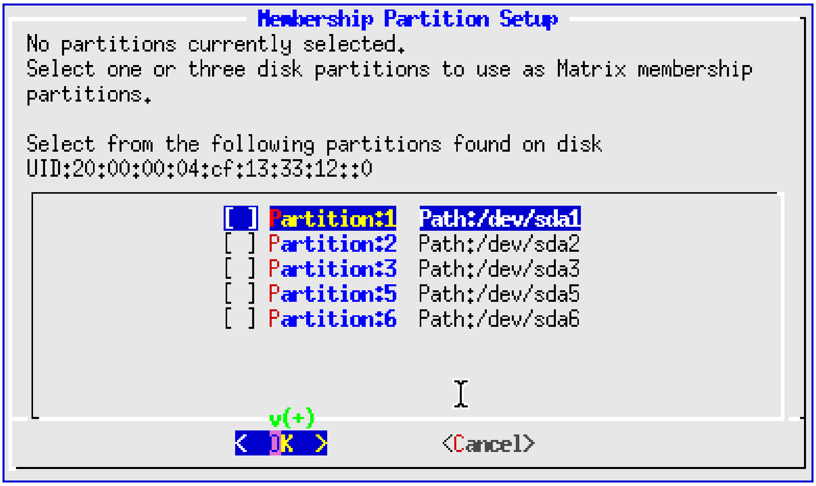
NOTE: If you want to add another membership partition or to remove or replace an existing membership partition, use the Repair option.

The Setup option asks you to select a SAN disk where you want to create a membership partition. Information about each disk appears at the bottom of the window. HP Clustered File System can operate with either one or three membership partitions. To ensure that a membership partition is always available, we recommend that you use three membership partitions.



```
Membership Partition Setup
No partitions currently selected.
Select one or three disk partitions to use as Matrix membership
partitions.

Select a disk to see a list of partitions on that disk.

    UID:20:00:00:04:cf:13:33:12::0    switch fcswitch3 port 7
    UID:20:00:00:04:cf:13:3c:92::0    switch fcswitch3 port 7
    Help                              Display help
```

When you select a disk, the partitions on that disk are displayed. To select a partition, move to that partition and press the spacebar. Information about the partition then appears at the bottom of the window. An 8-MB partition is adequate.



```
Membership Partition Setup
No partitions currently selected.
Select one or three disk partitions to use as Matrix membership
partitions.

Select from the following partitions found on disk
UID:20:00:00:04:cf:13:33:12::0

    [ ] Partition:1  Path:/dev/sda1
    [ ] Partition:2  Path:/dev/sda2
    [ ] Partition:3  Path:/dev/sda3
    [ ] Partition:5  Path:/dev/sda5
    [ ] Partition:6  Path:/dev/sda6
```

The partition you selected is displayed on the Membership Partition Setup window. If you want to use three membership partitions, repeat this procedure to select the additional membership partitions.

We recommend that the partitions be on different disks. You can deselect and reselect partitions as necessary. To complete the operation, select Done.

Administration When Cluster Is in a Degraded State

In certain rare error scenario combinations, it is possible for the cluster to be in a state where it cannot start up successfully because it believes that certain servers should not have access to the SAN; however, the cluster cannot exclude, or fence, the servers from the SAN because it cannot access one or more of the FC switches.

When the cluster is in this state, the cluster log will typically contain error messages in the following format:

```
CRITICAL ERROR: Unable to fence host xxxxxxxx xxxxxxx: xxxx
(switch=xxxx)
```

The FC switch that cannot be accessed is listed at the end of the error message.

To resolve the problem, bring the bad FC switch back online. If it is not possible to do this, stop the cluster and use **mxmpconf** to clear the host registry (see “Repair Menu Options” on page 187). Clearing the registry allows the cluster to start without fencing any servers.

CAUTION: Before clearing the server registry, be sure to reboot or power off any servers that were previously removed from the cluster and no longer had access to the SAN. After the servers have been rebooted, they can safely access the SAN. (If the servers are not rebooted, it is possible for them to corrupt filesystems.)

Changing the Number of LUNs Can Cause Cluster Failure

HP Clustered File System uses worldwide names and LUN numbers to associate PSD device names with SAN devices.

If the LUN numbers are changed on an array (for example, by combining LUNs or breaking a LUN into multiple LUNs), the cluster can fail to start when servers are rebooted because the PSD device names are no longer associated with the correct LUNs. HP Clustered File System cannot automatically detect a change in LUN numbering because its identification mechanism uses the worldwide name, which changes as the LUN changes.

To minimize downtime, if you need to make changes to your storage which will result in the LUN numbers on the array changing, contact HP Technical Support before making the change.

However, if this situation should occur, you can use the **mxmpconf** utility to restore the membership partitions. You will then be able to restart the cluster. For assistance in restoring access to PSFS filesystems, contact HP Technical Support.

NOTE: LUNs can safely be resized. Problems occur only if LUN numbering on the array is changed.

Server Cannot Be Located

If the cluster reports that it cannot locate a server on the SAN but you know that the server is connected, there may be an FC switch problem.

On a Brocade FC switch, log into the switch and verify that all F-Port and L-Port IDs specified in **switchshow** also appear in the local nameserver, **nsshow**. If the lists of ports are different, reboot the switch. If the reboot does not clear the problem, there may be a problem with the switch.

Also consult your FC switch documentation or the FC switch vendor. If the switch appears to be operating properly, contact HP Technical Support.

Online Insertion of New Storage

HP Clustered File System supports online insertion (OLI) of new storage, provided that OLI support is present for your combination of storage device, SAN fabric, HBA vendor-supplied device driver, and the associated HBA vendor-supplied libhbaapi. (Check with your vendors to determine whether OLI is supported.)

When this lower-level OLI support is in place, inserting a new disk will cause a new device to automatically become eligible for importing. The disk can be imported with the Management Console or **mx** utility, and can be used normally from that point forward.

If OLI is not possible with your hardware combination, you will need to restart HP Clustered File System after inserting a new disk. The disk will then be visible to the cluster.

Online Replacement of a FibreChannel Switch

When a cluster includes multiple FibreChannel switches, you can replace a switch without affecting normal cluster operations. The following conditions must be met when performing online replacement of a FibreChannel switch:

- The replacement switch must be the same model as the original switch and must have the same number of ports.
- The FC connectors must be reinserted in the same location on the new switch. For example, the FC connector that was plugged into port 1 on the original switch must be plugged into port 1 on the new switch.
- HP Clustered File System must be stopped on any servers that are connected only to the switch to be replaced.

If these conditions are not met, you will not be able to perform online replacement of the switch. Instead, you will need to stop the cluster, replace the switch, and use **mxconfig** to reconfigure the new switch into the cluster.

Consult your switch documentation for the appropriate replacement procedure, keeping in mind that the above requirements must be met. However, if this documentation is not available, the following procedures describe a method to replace a switch.

Replace a Brocade FC Switch

To replace a Brocade FibreChannel switch, complete these steps:

1. Stop HP Clustered File System on any servers that are connected only to the original switch.

2. If possible, save the configuration information on the original switch. Use the **configUpload** command.
3. Record the IP address of the original switch. Use the **ipAddrShow** command.
4. Back up the zone configuration information, either from the original switch or from another switch in the fabric. Use the **cfgShow** command and record its output.
5. Connect the power and either the Ethernet or the serial console cable to the new switch.
6. Log on to the new switch.
7. Disable the switch with the **switchDisable** command.
8. Disable any stale active configuration on the new switch with the **cfgDisable** command.
9. Verify that the Brocade licenses are installed by using the **licenseShow** command. The new switch should have the same kind of license as the rest of the fabric.
10. Clear any stale zone configuration on the new switch with the **cfgClear** command.
11. Save the clean configuration with the **cfgSave** command.
12. Configure the new switch. If you saved the original configuration with the **configUpload** command, use the **configDownload** command to restore it. Otherwise, use the **configure** command. (You may need to consult your site's SAN administrator or your Brocade representative for the correct configuration information.)
13. Connect the FC connectors to the new switch. Be sure to plug them into the same ports as on the original switch.
14. Set the Ethernet IP address on the new switch to the IP address of the original switch. Use the **ipAddrSet** command.

    If you had connected to the switch over the Ethernet interface, the session will be disconnected and you will need to log back into the switch with the IP address you just configured.

15. Enable the switch using the **switchEnable** command. The new switch should now connect to the rest of the fabric.
16. Verify that the HBAs on the servers log into the new switch in the expected locations. Also verify that no zone conflicts are being reported on the inter-switch links (ISL).
17. Verify that the new switch has been configured into the cluster. Run the **/opt/hpcfs/sbin/mxmpio status** command, which shows whether failover is enabled or disabled. (See **mxmpio**(8) for details.)
18. Verify that I/O operations are successful via the new switch. Mount a psd device, and then use **mxmpio** to set its active path to one of the paths that goes through the new switch. Then perform I/O operations such as creating or deleting files on the mounted psd device.

Replace a McDATA FC Switch

To replace a McDATA FibreChannel switch, complete the following steps.

NOTE: This procedure uses the McDATA Web interface to perform various FibreChannel switch operations. You can also use the McDATA command-line utility to perform these operations.

1. Stop HP Clustered File System on any servers that are connected only to the original switch.
2. If possible, save configuration information from the original switch. Some items such as the zone configuration are not needed and are just insurance against further failures. Other items such as the IP address are available elsewhere but might conveniently be captured here. One way to record the information is to capture the output of a CLI session.

   The following commands show types of data that might be useful:

   show ip ethernet for the IP address.

   show switch for the fabric operating mode.

   show zoning for the zone configuration.

3. After the original switch has been powered down, power up the new switch and set its IP address to address of the old switch to allow EWS access. The IP address can be set either from the serial port or by connecting to the current, possibly default, IP address. Verify that the new switch responds to the IP address.

4. Connect to the EWS and take the switch offline using the Operations > Switch > Online State tab.

5. Make the switch operating mode and domain ID acceptable to the original fabric. This can be done either by consulting the fabric or by taking the values from the data saved in step 2. This procedure might include changing the default zone setting as directed by EWS when changing interoperation mode. Any existing zone configuration on the new switch should be removed to allow the fabric to properly communicate current zoning when the switch joins the fabric.

6. Add the private community to the Configure > Management > SNMP tab and ensure that it is write enabled.

7. Connect the FC connectors to the new switch. Be sure to plug them into the same ports as on the original switch.

8. Bring the switch online. The new switch should now connect to the rest of the fabric. Verify that the HBAs on the servers log into the new switch in the expected locations.

9. Verify that the new switch has been configured into the cluster. Run the **/opt/hpcfs/sbin/mxmpio status** command, which shows whether failover is enabled or disabled. (See **mxmpio**(8) for details.)

10. Verify that I/O operations are successful via the new switch. Mount a psd device, and then use **mxmpio** to set its active path to one of the paths that goes through the new switch. Then perform I/O operations such as creating or deleting files on the mounted psd device.

16

Other Cluster Maintenance

Although HP Clustered File System requires little special maintenance beyond that which is normally required for your servers and services, you may need to perform the following activities:

- Maintain log files
- Disable a server for maintenance
- Troubleshoot a cluster
- Troubleshoot service and device monitors

Maintain Log Files

HP Clustered File System stores its log files in the */var/log/hpcfs* directory on each server in the cluster. These log files are typically used to record the actions of the HP Clustered File System daemons and agents. The *matrix.log* log file is of particular interest and is the log file for the ClusterPulse daemon. You can use the Management Console to view this log file and to manage its size.

The matrix.log File

The *matrix.log* file stores messages from the ClusterPulse daemon. You can also add your own user-defined messages to this file.

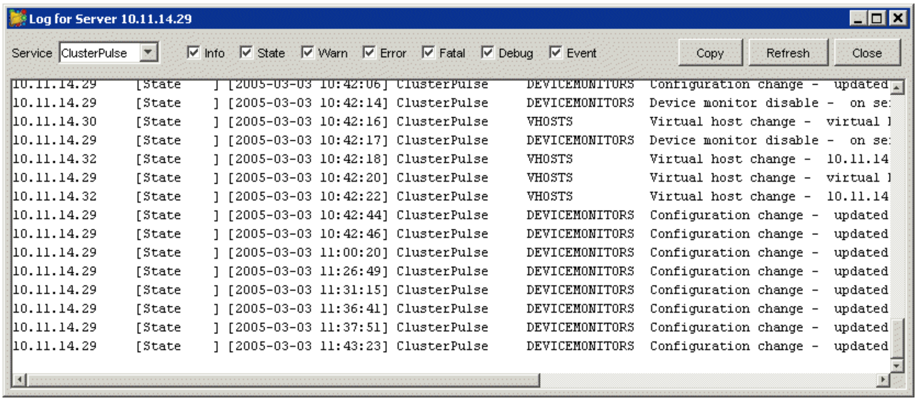
The messages in the *matrix.log* file are either local or global.

Local messages appear only in the *matrix.log* file on the server where the message originated. Global messages are distributed to each server in the cluster and are written into the *matrix.log* file on each server.

You can use the Management Console to view or maintain the *matrix.log* file on each server. The changes you make affect only the log file for the server selected on the Servers window.

Read the Cluster Log File

Select the server where you want to view the log, right-click, and then select View Log.

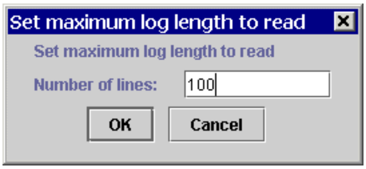


```
10.11.14.29     [State    ] [2005-03-03 10:42:06] ClusterPulse     DEVICEMONITORS  Configuration change -  updated
10.11.14.29     [State    ] [2005-03-03 10:42:14] ClusterPulse     DEVICEMONITORS  Device monitor disable -  on se:
10.11.14.30     [State    ] [2005-03-03 10:42:16] ClusterPulse     VHOSTS          Virtual host change -  virtual ]
10.11.14.29     [State    ] [2005-03-03 10:42:17] ClusterPulse     DEVICEMONITORS  Device monitor disable -  on se:
10.11.14.32     [State    ] [2005-03-03 10:42:18] ClusterPulse     VHOSTS          Virtual host change -  10.11.14
10.11.14.29     [State    ] [2005-03-03 10:42:20] ClusterPulse     VHOSTS          Virtual host change -  virtual ]
10.11.14.32     [State    ] [2005-03-03 10:42:22] ClusterPulse     VHOSTS          Virtual host change -  10.11.14
10.11.14.29     [State    ] [2005-03-03 10:42:44] ClusterPulse     DEVICEMONITORS  Configuration change -  updated
10.11.14.29     [State    ] [2005-03-03 10:42:46] ClusterPulse     DEVICEMONITORS  Configuration change -  updated
10.11.14.29     [State    ] [2005-03-03 11:00:20] ClusterPulse     DEVICEMONITORS  Configuration change -  updated
10.11.14.29     [State    ] [2005-03-03 11:26:49] ClusterPulse     DEVICEMONITORS  Configuration change -  updated
10.11.14.29     [State    ] [2005-03-03 11:31:15] ClusterPulse     DEVICEMONITORS  Configuration change -  updated
10.11.14.29     [State    ] [2005-03-03 11:36:41] ClusterPulse     DEVICEMONITORS  Configuration change -  updated
10.11.14.29     [State    ] [2005-03-03 11:37:51] ClusterPulse     DEVICEMONITORS  Configuration change -  updated
10.11.14.29     [State    ] [2005-03-03 11:43:23] ClusterPulse     DEVICEMONITORS  Configuration change -  updated
```

The Server Log window displays the most recent messages from the *matrix.log* file. You can select the types of messages that you want to view by checking or unchecking the boxes at the top of the window. Use the scroll bars to move up, down, left, and right in the file, allowing you to see entire messages without resizing the window.

Set the Number of Lines to Display on the Server Log Window

Select the server on the Servers window, right-click, and select Set Log Length. Then specify the maximum number of lines that will be displayed with the View Log option. The default is 100 lines. This option has no effect on the *matrix.log* itself.


Set maximum log length to read

Set maximum log length to read

Number of lines: 100

Rotate the Cluster Log File

HP Clustered File System rotates the cluster log file on a regular basis. Up to five old versions of the file are saved in the */var/log/hpcfs* directory; the saved files are named *matrix.log.1*, *matrix.log.2*, and so on. You can also rotate the cluster log file from the Management Console. Select the appropriate server on the Servers window, right-click, and select Rotate Log. The Management Console will display an error message if the rotate option fails.

Add Your Own Messages to the Cluster Log File

You can use the **mxlogger** command to add your own messages to the *matrix.log* file. For example, you may want to add messages about the state of your applications. The **mxlogger** command has this syntax:

```
mxlogger -e entity -l level [-G|-L] log-text
```

The **-e** *entity* option specifies the type of message that you are adding. *entity* can be USER1, USER2, USER3, USER4, USER5, USER6, USER7, or USER8. You will need to determine how you want to use these entities.

The **-l** *level* option specifies the severity of the message. *level* can be ERROR, WARN, INFO, EVENT, FATAL, STATE, TRACE, or DEBUG.

The **-G** option specifies that the message is global; the **-L** option specifies that it is local. The default is local.

If the *log-text* contains special characters or spaces, it must be enclosed in quotation marks.

The following command adds a local message to the *matrix.log* file:

```
mxlogger -l info -e User2 “hello, world.”
```

The message appears like this in the log file:

```
Server      Level  Date/time             Facility Entity Message
192.168.0.1 [Info] [2001-10-07 14:16:27] User     USER2  hello, world
```

Cluster Alerts

The Alerts section at the bottom of the Management Console window lists errors that have occurred in cluster operations. Double click an alert to view the error in the cluster tree structure.

The **mx alert status** command can also be used to display the current alerts. See the *HP StorageWorks Clustered File System Command Reference Guide* for details about this command.

Disable a Server for Maintenance

If you need to perform maintenance (such as an application software upgrade) on a server that is providing the primary interface for a virtual host, take these steps before working on the server:

1. Disable the server. (Choose the server from the Servers window, right-click, and select Disable.) This step causes the virtual host to fail over to a backup network interface on another server.
2. If you want the virtual host to remain on the backup network interface after the original server is returned to operation, make that network interface the primary network interface. (Choose the virtual host from the Virtual Hosts window, right-click, and select Properties.)
3. Perform the necessary maintenance on the original server and then reenable it.

Detection of Down Servers

The ClusterPulse daemon uses heartbeats to determine whether a server is up. At a specific interval, ClusterPulse sends a heartbeat message to each server. This is called a "heartbeat event." Each server is then required to send a response back to ClusterPulse.

The suspect interval specifies the number of heartbeat events that can pass without receiving a response from a server. If the server does not respond within this interval, ClusterPulse determines that the server is down.

The default value for the suspect interval is 34. If your cluster is focused on SAN activities and servers are being reported as down during times of high load average and/or high disk utilization, it may be useful to increase this value. If your cluster is focused on services, such as HTTP or FTP, you may want to decrease the value.

NOTE: Changing the suspect value can impact failure detection. If the suspect interval is too high, ClusterPulse may not immediately detect that a server is down. If the interval is too low and a server does not have enough time to respond to a heartbeat, ClusterPulse may incorrectly determine that the server is down.

To change the suspect interval, add a "clusterpulse_start_options" line such as the following to the */etc/hpcfs/mxinit.conf* file, where *n* is the new value.

```
clusterpulse_start_options = { "-nodaemon", "-suspect n" };
```

Troubleshoot Cluster Problems

The following situations do not produce specific error messages.

The Server Status Is "Down"

If a server is running but HP Clustered File System shows it as down, follow these diagnostic steps:

1. Verify that the server is connected to the network.
2. Verify that the network devices and interfaces are properly configured on the server.
3. Ensure that the ClusterPulse daemon is running on the server.
4. Verify that the same version of HP Clustered File System is installed on all servers in the cluster.

A Running Service Is Considered Down

If HP Clustered File System indicates that a service is down even though it is running, the service might not be sending any information to HP Clustered File System. HP Clustered File System will conclude that the service is down even if the TCP connection succeeds. In particular, if an HTTP server is monitored by a FTP monitor, the HTTP server is considered down.

Also check the following:

1. Verify that the server is connected to the network.
2. Verify that the network devices and interfaces are properly configured on the server.
3. Ensure that the ClusterPulse daemon and the service monitor agent (**monitor_agent**) are running on the server.

A Virtual Host Is Inaccessible

If a site is inaccessible but HP Clustered File System indicates that it is okay, verify that the service is running and that the actual data exists on the server providing the network interface currently used by the virtual host.

HP Clustered File System Exits Immediately

If the ClusterPulse daemon exits immediately on starting, check the last lines of the following files for errors:

- */var/log/hpcfs/matrix.log*
- */var/log/hpcfs/mxinit.log*

This problem typically occurs because either the hostname is not set properly on the server or the main Ethernet interface is not installed.

Refer to the **ifconfig** man page for ways to check this.

Browser Accesses Inactive Server

If you access a virtual host from a Web browser that is running on one of the servers in a cluster, the browser accesses the service on the server it is running on. Note that this allows you to monitor and test the inactive server while the virtual host is fully operational.

Troubleshoot Monitor Problems

You may encounter the following problems with service and device monitors.

Monitor Status

If the monitor status is not reported as Up, check the last error message string and the last event message string that **monitor_agent** returned to HP Clustered File System for any service or device monitor on any server in the cluster. The error or event message provides more status information.

"Down" Status

The "Down" status indicates that the monitor finished its probe but it did not complete successfully. Depending on the monitor type (such as HTTP or SMTP), the service monitor probe may involve more than being able to connect to the network service.

For many built-in service monitors, HP Clustered File System may conclude that the monitor is down even if the TCP connection succeeds. For example, if a Web server is available but the URL specified in the HTTP service monitor was not found, the HTTP service monitor will be reported as Down.

"Undefined" Status

If the probe has not completed because of a script configuration problem or because HP Clustered File System is still attempting to finish the first probe, the status will be reported as "undefined" instead of Down.

“SYSTEM ERROR” Status

The “SYSTEM ERROR” status indicates that a serious system functional error occurred while HP Clustered File System was trying to probe the service. To see details about the system error, select the monitor on the Management Console, right-click, and select View Last Error. Also view the *matrix.log* file.

“Unknown” Status

The “unknown” status indicates that the information that ClusterPulse provided to the Management Console did not conform to one of the expected status choices. This could occur if the Management Console is out of date and does not support the version of HP Clustered File System running on the server.

“Event” Status

The “Event” status is displayed when **monitor_agent** encounters an error while executing the probe, Start, Stop, or Recovery scripts.

The status of the monitor may be “Up” even though an event has been reported. This is the case if the error was not serious enough to stop the execution of the script.

The events are:

CONFIG_ERROR. A script must exist and be executable by *root*. This condition is checked for probe, Start, Stop, and Recovery scripts each time an attempt is made to execute the script.

SCRIPT_SYSERR. The **monitor_agent** tried to fork a process to execute the script but the fork system call failed. This condition can occur when HP Clustered File System is trying to execute probe, Start, Stop, or Recovery scripts.

START_TIMEOUT. A Start script was executed but it did not complete within the specified timeout period.

STOP_TIMEOUT. A Stop script was executed but it did not complete within the specified timeout period.

RECOV_TIMEOUT. A Recovery script was executed but it did not complete within the specified timeout period.

START_FAILURE. A Start script was executed but it returned a non-zero exit status.

STOP_FAILURE. A Stop script was executed but returned a non-zero exit status.

RECOV_FAILURE. A Recovery script was executed but returned a non-zero exit status.

ILLEGAL TRANSITION. The ClusterPulse daemon gave instruction to the **monitor_agent** daemon, which is considered to be an illegal state transition for a monitor. This indicates an internal error and should be reported to HP Technical Support.

The event is written into the cluster log. To view the error, select the monitor on the Management Console, right-click, and select View Last Error. You can also view the information in the *matrix.log* file.

From the command line, use these commands to check the status of monitors:

```
mx service status
mx device status
```

Clear an Error

After you have determined the cause for a script event, be sure to correct the script on all servers that reported the event. You should then clear the error. On the Management Console, select the monitor, right-click, and select Clear Last Error. To clear an error from the command line, use these commands:

```
mx service clear <vhost:port> <server> ...
mx device clear <devicename> <server> ...
```

Because the error is server-specific, you must clear it on each server in the cluster (just as you had to correct the script on each server that reported a problem).

NOTE: An error on a monitor may still be indicated after correcting the problem with the Start, Stop, Recovery, or probe script.

Errors can be cleared only with the Management Console or the appropriate **mx** command. An error will not be automatically cleared by the ClusterPulse daemon. The date and time of the script event are available in the *matrix.log* file.

Monitor Activity

Service and device monitors report one of the following as their activity status:

- (activity unknown)
- Transitioning
- Starting
- Stopping
- Inactive
- Active

The activity status is not an error condition; it represents the activity of scripts associated with the monitor. However, if the activity status continues to have a value other than Active or Inactive, there may be a script problem that requires attention. Active status indicates that the probe script will be executed at the probe frequency. Inactive status means that the monitor is not the only one currently providing the service. For example, when the primary server is functioning normally, a monitor on a backup server may show an Inactive status. The probe script is still executed in the Inactive state.

Typically, the status is Active on all appropriate servers, as it is important to continue to probe the service on backup servers as well as on the primary servers.

“Activity Unknown” Status

For a brief period while the **monitor_agent** daemon checks the monitor script configuration and creates a thread to serve the monitor, the activity may be displayed as “activity unknown.”

“Transitioning” Activity

The “Transitioning” activity indicates that the monitor state is on its way to becoming ACTIVE or INACTIVE (or starting or stopping, if a Start or Stop script is present).

“Starting” Activity

This activity is reported only when a Start script has been specified for the monitor. It indicates that the **monitor_agent** thread is running the Start script and the monitor is in the STARTING state.

“Stopping” Activity

This activity is reported only when a Stop script has been specified for the monitor. It indicates that the **monitor_agent** thread is running the Stop script and the monitor is in the STOPPING state.

Service Monitor Always Down

A network service monitor such as a generic TCP port monitor will always appear as “Down” if the network application is configured to bind to the host address and not to the virtual host address. One solution is to modify the application so that it binds with the address of the virtual host or the value of INADDR_ANY. Remember, service monitors probe the virtual host address and not the server’s host address. You can use the **netstat** command to display the listening server sockets. If the Local Address appears as *host:port* instead of *virtual-host:port* or **:port*, the server software should be reconfigured as described above.

Monitor Recovery

If a Recovery script is configured, a probe failure will not immediately cause the monitor to become “Down.” The Recovery script is attempted only once after the probe fails. However, when the monitor again reaches the “Up” status, the Recovery script is again enabled to run if the probe fails again later. A Recovery script is a valuable tool if you expect the service or device to fail periodically and you do not want to take the failover action for a single probe failure. Putting a script like this in place essentially implements a “two consecutive probe-script failure” probe.

Firewall Information

HP Clustered File System uses these ports for server-to-server communication.

UDP Ports	TCP Ports
7659	9050
7660	9060
7661	9070
8940	9071
9046	
23646	

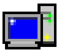
Communication between the administration console and the server is handled via TCP port 9050.

A

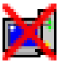
Management Console Icons

The Management Console uses the following icons to represent the status of HP Clustered File System entities.

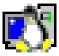
Servers



The server is up.

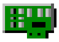


The server is down.

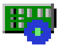


The server icon can also contain an image indicating the operating system in effect on the server (for example, Linux).

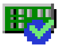
Network Interfaces



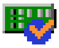
The network interface discourages administrative traffic.



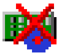
The network interface allows administrative traffic.



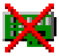
The network interface is currently the administrative network and was configured to allow administrative traffic.

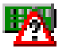


The network interface is currently the administrative network but was configured to discourage administrative traffic. No other interface could be used by all of the servers.



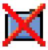
A network interface that allows administrative traffic is down.


A network interface that discourages administrative traffic is down.

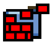

The netmask for this network interface is unknown.

PSFS Filesystems

The filesystem is active.

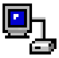

The filesystem is down.

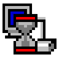
The filesystem mount is persistent.

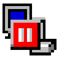

The filesystem is undergoing recovery.

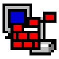
The filesystem has been recovered.

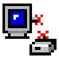
PSFS Filesystem Mounts

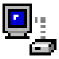

The filesystem is mounted.

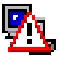

The filesystem is being mounted.


The mount operation has been paused.

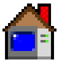

The mount operation is recovering.

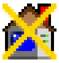

The filesystem is being unmounted.


The filesystem is unmounted

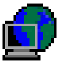

The filesystem has experienced an error.

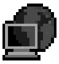
Virtual Hosts

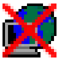

The virtual host is active on the primary server, or is inactive but available on a backup server.

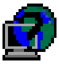

The virtual host is down.

Service Monitors

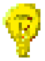
 The service monitor is enabled.

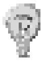
 The service monitor is disabled.

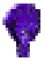
 The service monitor is down.

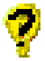
 The state of the monitor is unknown.

Device Monitors

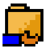
 The device monitor is enabled.

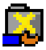
 The device monitor is disabled.

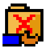
 The device monitor is down.

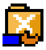
 The state of the monitor is unknown.

Export Groups

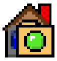
 The export group device monitor is enabled.

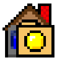
 The export group device monitor is disabled.

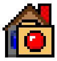
 The export group device monitor is down.

 The state of the monitor is unknown.

Virtual NFS Services

 The Virtual NFS Service is active.

 The Virtual NFS Service is inactive.

 The Virtual NFS Service is down.

Applications Tab

The green arrow specifies that the HA application monitor is active on this server. In this example, "P" indicates that this is the primary interface.

The grey bars indicate that the HA application monitor is inactive on this server. In this example, "1" indicates that this is the first backup.

The red bars indicate that the HA application monitor is disabled on this server. In this example, "2" indicates that this is the second backup

The yellow bars indicate that the HA application monitor is transitioning to another state.

The red square specifies that the HA application monitor is down.

The green arrow specifies that the scale-out application monitor is up. This status applies to the entire cluster.

The red arrow specifies that the scale-out application monitor is down. This status applies to the entire cluster.

B

Error and Log File Messages

When certain errors occur, HP Clustered File System writes messages to the Management Console. Other error messages are written to the server's log file (*matrix.log*).

Management Console Alert Messages

```
NN.NN.NN.NN has lost a significant portion of its SAN
access, possibly due to a SAN hardware failure
```

The specified server is unable to write to any of the membership partitions. Ensure that the server can access the membership partitions and also has write access to them. Also check for hardware problems that can limit access to the partitions. If the disks containing the membership partitions have failed, you will need to recreate the partitions. (See "Manage Membership Partitions with mxmpconf" on page 183 for more information.)

```
NN.NN.NN.NN is unresponsive and must be rebooted ASAP.
Until this reboot is done, some or all filesystem operations
may be paused throughout the cluster.
```

The specified server did not respond to a critical HP Clustered File System request. This problem may be caused by a severe overload or resource shortage on the server. Rebooting the server should clear the problem.

```
NN.NN.NN.NN should be rebooted ASAP as it stopped cluster
network communication DATE HH:MM:SS and was excluded from
the SAN to protect filesystem integrity
```

The server was excluded from the cluster because it could no longer communicate over the network. The server should be rebooted at the first opportunity. Also check the network and make sure that the server is not experiencing a resource shortage.

```
NN.NN.NN.NN should be rebooted ASAP as it was excluded from
the SAN to protect filesystem integrity
```

The server was excluded from the cluster because it could no longer communicate over the network. The server should be rebooted at the first opportunity. Also check the network and make sure that the server is not experiencing a resource shortage.

```
Error adding <entity>: I/O error sending to server

Error disabling <entity>: I/O error sending to server

Error enabling <entity>: I/O error sending to server

Error updating <entity>: I/O error sending to server
```

If you receive one of these errors, you can attempt to resolve it by restarting HP Clustered File System. If this is unsuccessful, call HP Technical Support.

```
Error connecting to server <host>: unknown host
```

The server identified as *<host>* is not responding to the connection request from the Management Console. Verify that you typed a valid hostname or IP address in the Connect to Cluster window. This error may indicate that the ClusterPulse daemon is not running; restart HP Clustered File System on the server.

```
Error connecting to server: I/O error
```

```
Error connecting to server
```

An error occurred when HP Clustered File System tried to connect to the specified server through the Connect to Cluster window.

```
Error getting cluster status from server: <exception string>
```

The <*exception string*> describes the error. The connection to the server on port 9050 was successful but the first response from the server experienced an I/O error.

```
Error logging in: I/O error connecting to server
```

An I/O error occurred during the authentication of the Management Console to the ClusterPulse daemon. View recent entries in the *matrix.log* file for indications of what might have happened.

```
Error when connecting to the server: version not found -
Using the latest version in the cache
```

By default, the Management Console checks the current software version on the server to which it is being connected and then downloads the appropriate software only if that version is not already in the local cache. This message indicates that the version check failed and the Management Console is using the software in its local cache.

```
Fencing operation failed, reboot NN.NN.NN.NN ASAP.
NN.NN.NN.NN stopped cluster communication DATE HH:MM but
cannot be excluded from the cluster because of a networking
or fencing hardware failure or misconfiguration. To protect
filesystem integrity, some or all filesystem operations may
be paused until NN.NN.NN.NN is rebooted or until fencing
operations can be performed.
```

The server is unresponsive but HP Clustered File System is unable to exclude it from the cluster because the fencing device is not responding to the fencing request. Possibly the Ethernet connection to the FibreChannel switch has failed. The server should be rebooted as soon as possible.

```
Internal error: unable to initialize security
```

```
Internal error: unable to initialize security. Java program
problem. Contact service.
```

If you receive one of these messages, report it to HP Technical Support at your earliest opportunity.

```
Majority of membership partitions are unwritable, possibly
due to a SAN or storage hardware failure. As a result, disk
imports and deports cannot be done, and some servers may be
unable to mount filesystems. In addition, HP Clustered File System's
ability to recover from a future server failure is
compromised. Such a failure would leave HP Clustered File System no
option but to pause some or all filesystems throughout the
cluster to preserve filesystem integrity.
```

The servers in the cluster cannot write to the majority of the membership partitions. Ensure that the servers can access the membership partitions and have write access to them. If the disks containing the membership partitions have failed, you will need to recreate the partitions. (See "Manage Membership Partitions with mxmpconf" on page 183 for more information.)

```
Cluster unable to take control of SAN, because a majority of
the membership partitions cannot be written or are corrupt,
possibly due to a SAN hardware failure or misconfiguration
and/or because servers have been excluded from the SAN. As a
result, some or all filesystem operations may be paused
throughout the cluster. In addition, filesystem mounts and
unmounts and disk imports and deports cannot be performed.
```

Ensure that all servers can access the membership partitions and have write access to them. If a membership partition is corrupt, another alert will specify its location. See "Manage Membership Partitions with mxmpconf" on page 183 for corrective actions.

```
Cluster unable to take control of SAN, because another cluster
that includes NN.NN.NN.NN currently controls the SAN.
Possibly a networking failure or misconfiguration has
partitioned these servers from the servers that control the
SAN, or possibly this cluster has been misconfigured to
share membership partitions with another cluster.
```

Check the cluster configuration and add the server if it is not currently a member. If your site has more than one cluster, use **mxmpconf** to make sure that this server is using the correct set of membership partitions.

```
Cluster unable to take control of SAN, because the servers are
unable to perform fencing operations, possibly due to a
networking or fencing hardware failure or misconfiguration.
As a result, some or all filesystem operations may be paused
throughout the cluster. In addition, filesystem mounts and
unmounts and disk imports and deports can not be performed.
```

HP Clustered File System is unable to exclude servers that have become unresponsive, or to verify the fencing of servers that were previously fenced and have not yet been rebooted. Possibly the fencing device is not responding to the fencing request or the Ethernet connection to the FibreChannel switch has failed. Also check for other networking or SAN hardware problems. The servers should be rebooted as soon as possible.

```
Cluster unable to take control of SAN, because the servers are
unable to write to a majority of the membership partitions,
possibly due to a SAN hardware failure or misconfiguration
and/or because some servers have been excluded from the SAN.
As a result, some or all filesystem operations may be paused
throughout the cluster. In addition, filesystem mounts and
unmounts and disk imports and deports cannot be performed.
```

Check the cluster configuration and add the server if it is not currently a member. Also ensure that the server can access the membership partitions and has write access to them, and check for any SAN hardware problems.

```
Cluster unable to take control of SAN. As a result, some or
all filesystem operations may be paused throughout the
cluster. In addition, filesystem mounts and unmounts and
disk imports and deports cannot be performed.
```

Verify that this message, not one of the previous messages, is displayed. Also verify that the servers can access the membership partitions and have write access to them, and that the servers can communicate with the FibreChannel switch. If no problems are found, contact HP Technical Support.

```
Membership partition XXXX is unwritable, possibly due to a
SAN or storage hardware failure. If other membership
partitions become inaccessible, HP Clustered File System's ability
to recover from a server failure will be compromised.
```

None of the servers in the cluster can write to the specified membership partition. Ensure that the servers can access the membership partition and that they have write access for it. If the partition is corrupt, another alert will display the location of the partition. See "Manage Membership Partitions with mxmpconf" on page 183 for corrective actions.

```
Membership Partitions are corrupt or inaccessible,
preventing SAN access
```

A majority of the membership partitions are either inaccessible or corrupt. HP Clustered File System cannot allow access to the PSFS filesystems while the Membership Partitions are in this state. To obtain more detailed information about the state of each Membership Partition, use the **mxmpconf** command.

If membership partitions are reported as not found or inaccessible, there is a problem with accessing the disk devices; check your hardware configuration. If membership partitions are reported as corrupt, then **mxmpconf** can be used to repair them, as long as at least one membership partition is reported to be in a good (OK) state. (See "Manage Membership Partitions with mxmpconf" on page 183 for more information.)

```
psdNpN on NN.NN.NN.NN is stalled on locks from NN.NN.NN.NN
```

A DLM lock request has been outstanding for a long period of time on the specified server. Probably the server is severely overloaded or is experiencing a resource shortage. As a last resort, reboot the server to clear the problem.

```
Server closed connection
```

```
connection to the server was lost
```

Either the server has closed the communication connection to the Management Console or the ClusterPulse daemon has died on the server you were logged onto.

Log on to the server if you are remote. Verify that the ClusterPulse daemon is still executing and restart HP Clustered File System if necessary.

```
Server protocol version received from <host> is incompatible
with this console version.
```

Verify that you are using the version of the Management Console provided with HP Clustered File System.

```
Singleton cluster unable to take control of SAN, because the
cluster that includes NN.NN.NN.NN currently controls the SAN.
Possibly this server has not been added to the cluster or has
been deleted from the cluster, or possibly a networking
failure or misconfiguration has partitioned this server from
the servers that control the SAN.
```

Check the cluster configuration and add the server if it is not currently a member. Also check the network to ensure that the server can communicate with the rest of the cluster.

```
Singleton cluster unable to take control of SAN. Possibly
this server has not been added to the cluster or has been
deleted from the cluster, or possibly a network failure has
partitioned this server from the rest of the cluster. As a
result, some or all filesystem operations may be paused
throughout the cluster. In addition, filesystem mounts and
unmounts and disk imports and deports cannot be performed.
```

Check the cluster configuration and add the server if it is not currently a member. Also check the network connections and configuration for the server to ensure that it can communicate with the rest of the cluster.

```
This monitor does not appear to be living on any particular
server. This is a UI bug.
```

If you receive this message, report it to HP Technical Support at your earliest opportunity.

```
Unknown user - NO_SUCH_USER

Unknown user - SYSTEM_ERROR

Unknown user - SYSTEM_ERROR: Unable to open pmxs_passwd
file.
```

The Management Console experienced an error while trying to authenticate itself to ClusterPulse. The *pmxs_passwd* file should be regenerated with the **mxpass** command.

ClusterPulse Messages

```
Bad command -- Could not find device monitor instance for
XXX on server YYY
```

The **monitor_agent** daemon is reporting status on a device monitor with device name *XXX* on server *YYY* but the ClusterPulse daemon does not recognize this device. Probably the Management Console has removed the device monitor and **monitor_agent** has already sent the status to ClusterPulse. Therefore, no corrective action is required.

```
Bad command -- HP Clustered File System daemon received an unknown
action or context reply from the agent monitor_agent:
[STATUS, EVENT,UNDEF] [SERVICE, DEVICE, UNKNOWN]
```

The message from the agent was not interpreted. This may indicate that the versions of the **monitor_agent** and ClusterPulse daemons are not the same.

```
Configuration error -- could not write to configuration file
```

The ClusterPulse daemon was unable to write to the specified configuration file. The text following the error message provides more information about the problem. For example, it may state that a filesystem is full.

```
Configuration file error -- could not open configuration
file
```

The ClusterPulse daemon was unable to open the configuration file */etc/hpcfs/cp_conf*. Verify that *root* has write access to the file.

```
Connection error -- send failed: N
```

```
Connection error -- transmit failed, no connection to
destination
```

```
Connection error -- total send failure
```

The ClusterPulse daemon experienced communication errors while trying to send status to the Management Console or an **mx** command. The connection to the UI process needs to be re-established.

```
Internal system error -- Internal error at server X.X.X.X:
select returned with an unknown read socket N
```

```
Internal system error -- Internal error at server X.X.X.X:
select returned with an unknown write socket N
```

```
Internal system error -- Internal select error at server
X.X.X.X: [select ?] with errno of N
```

The ClusterPulse daemon received a system error. Report this error to HP Technical Support at your earliest opportunity.

```
License error -- LICENSE ERROR ON SERVER %s: %s;
clusterpulse WILL BE TERMINATED IN %d HOURS %d MINUTES
```

The ClusterPulse daemon has recognized a license violation. This message will be repeated every 15 minutes. After two hours, HP Clustered File System will time out and ClusterPulse will terminate. HP Clustered File System can be started again, but this process will continue. Contact HP Technical Support to obtain a valid license or to install your license code correctly.

```
License error - LICENSE VIOLATION: <feature> is not licensed
on all servers.
```

The specified feature is not licensed on all servers. <*feature*> can be any licensed feature. The cluster will continue to operate properly.

```
License error - LICENSE VIOLATION: Cluster consists of
<cur_count> nodes with a maximum node count of <max_lic>
which exceeds the current license.
```

<*cur_count*> specifies the number of servers currently configured in the cluster and <*max_lic*> is the maximum number of nodes specified in any of the servers' license file. For example, servers Sa and Sb could have a cluster license for two nodes, while servers Sw, Sx, Sy, and Sz have a cluster license for four nodes. If you attempted to form a cluster with all of these servers, the <*cur_count*> would be six, and the <*max_lic*> would be four.

The cluster will continue to operate properly.

```
Monitor error -- monitor_agent reported N:: <error string>
```

The **monitor_agent** daemon experienced an error and is copying the error string to the *matrix.log* file. Inspect the error string for details about resolving the error.

```
Network error -- set_readable called with unknown socket N
```

```
Network error -- set_writeable called with unknown socket N
```

If you receive this message, notify HP Technical Support at your earliest convenience.

```
Object not found -- could not find service monitor
instance: IP X.X.X.X port N
```

The ClusterPulse daemon received a status message from the **monitor_agent** daemon regarding the service monitor for virtual host address IP *X.X.X.X* on port *N*, but the daemon does not recognize the service monitor. Probably the Management Console has removed the service monitor and **monitor_agent** has already sent the status to ClusterPulse. Therefore, no corrective action is required.

```
Read error - could not read configuration file
```

The ClusterPulse daemon experienced an error when reading from the specified configuration file */etc/hpcfs/cp_conf.* Verify that *root* has read access to the file.

```
Script error -- Internal consistency error in HP Clustered File
System
daemon: wrong parameters
```

If you receive this message, notify HP Technical Support at your earliest convenience.

```
Script error -- max pipe write retry attempt limit for
monitor agent monitor_agent. Shutting down agent.
```

```
Script error -- socket write error for agent monitor_agent.
Shutting down agent.
```

```
Script error -- Write to monitor failed: <errorno
description> for agent monitor_agent. Shutting down agent.
```

```
Script error -- Write to monitor failed: <errorno
description>. This probably means the agent has crashed for
agent monitor_agent. Shutting down agent.
```

The ClusterPulse daemon experienced an error while trying to write to the **monitor_agent** daemon. It will attempt to recover from this failure.

```
Script error -- HP Clustered File System cannot invoke a non executable
agent monitor_agent
```

Verify that the execute permission on **monitor_agent** is set correctly.

```
Script error -- HP Clustered File System daemon received an illegal
reply from the agent monitor_agent: <reply buf>
```

The ClusterPulse daemon received a reply buffer from the **monitor_agent** daemon that could not be parsed. Check the reply buffer for uppercase strings. If the strings are part of a script name, replace the string with some other string.

```
Script error -- HP Clustered File System failed to open stderr
sockets to agent monitor_agent
Script error -- HP Clustered File System failed to open stdout
sockets to agent monitor_agent
Script error -- HP Clustered File System failed to open stdin
sockets
to agent monitor_agent
```

An attempt to create a FIFO has failed because system resources are at their limits.

```
Script error -- HP Clustered File System spawn a monitor process failed
monitor_agent on server X.X.X.X in order to write <buffer>
```

```
Script error -- HP Clustered File System failed to launch
agent monitor_agent
```

An attempt to create a process has failed because system resources are at their limits.

```
Write error - in default_write_fun: Unknown connection mode
for IP %s port %d
```

```
Read error - in default_read_fun: Unknown connection mode
for IP %s port %d
```

If you receive either of these messages, notify HP Technical Support at your earliest convenience.

PSFS Filesystem Messages

If you receive a panic message from the PSFS filesystem, report it to HP Technical Support at your earliest convenience. Then reboot the affected server to recover from the error condition.

Distributed Lock Manager Messages

The Distributed Lock Manager (DLM) generates error messages if it detects that a filesystem operation will block indefinitely because of an internal error. The DLM also generates error messages if the **dlm** process is failing. If you receive one of these errors, report it to HP Technical Support at your earliest convenience.

The error messages have the following format:

```
time/date stamp:error:message text
```

Typically, you will need to reboot the affected server to recover from the error condition.

SANPulse Error Messages

SANPulse sends two types of messages to the *matrix.log* file: error and fatal.

Error messages have the following format:

```
[Error  ] [<date and time>] SANPulse SERVERS Internal error: ...
```

These messages indicate that an internal error has occurred in the SANPulse component. If you receive an error message, report it to HP Technical Support at your earliest convenience.

Fatal messages have this format:

```
[Fatal  ] [<date and time>] SANPulse      SERVERS
```

A fatal message indicates that HP Clustered File System has terminated on the specified server. First attempt to restart HP Clustered File System on that server. If the cluster software cannot be restarted, you will see another message asking you to reboot the server. The message will remain in the Alert section of the Management Console until either HP Clustered File System is restarted on the rebooted server or the server is removed from the cluster.

SCL Daemon Messages

If messages such as the following appear in the cluster log, the cluster may not be able to start up properly. See "Administration When Cluster Is in a Degraded State" on page 197 for a description of the problem and how to resolve it.

```
CRITICAL ERROR: Unable to fence host xxxxxxxx xxxxxxx: xxxx
(switch=xxxx)
```

Network Interface Messages

The PanPulse daemon generates messages about the state of the network interfaces configured in the cluster. If you receive a message that a network interface is down or unavailable, or that a network interface does not meet the interface requirements, you will need to take corrective actions.

A Network Interface Is Down or Unavailable

These messages indicate that there is a problem with the specified network interfaces. The hardware may have failed, or a network interface may have been taken down administratively. Check the status of the network interface and repair it if necessary.

If PanPulse determines that one or more local interfaces are down or unavailable, it will report messages such as the following:

```
Interface <ethn> address <xxx.xxx.xxx.xxx> has gone down

Interface <ethn> has gone down after a send failure

Interface <ethn> has gone down after an attempt to assign an
outbound multicast interface: <No such device>

Interface <eth_n> has been marked down because its interface
flags do not include IFF_UP
```

If PanPulse determines that the interfaces are down or unavailable on another server, it will report the following:

```
No interfaces are responding on host <xxx.xxx.xxx.xxx>
```

PanPulse should receive cluster traffic at certain intervals on the active administrative network interface. If PanPulse does not receive any traffic during this period of time, it will report the following message:

```
No traffic received on active panpulse interface (<ethn>)
```

PanPulse then fails over the active interface to another network interface and reports the following. (You will also see this message for other situations in which PanPulse chooses a new active interface.)

```
Selected new active interface <ethn> address
<xxx.xxx.xxx.xxx>
```

If PanPulse cannot communicate with all other active servers over a new interface, it will report the following message:

```
Unable to fail over to another interface
```

Network Interface Requirements Are Not Met

PanPulse generates error messages when the requirements for a network interface are not met.

Flags

HP Clustered File System requires that the following flags be set on network interfaces used in a cluster: UP, RUNNING, BROADCAST, MULTICAST. If any of these flags are not set, you will see one of the following messages. The **ifconfig**(8) man page describes how to set the required flags.

```
panpulse cannot use interface <ethn> because the required
interface flags (UP, RUNNING, BROADCAST, MULTICAST) are not
set. See ifconfig(8).
```

```
panpulse will not provide service until the required
interface flags (UP, RUNNING, BROADCAST, MULTICAST) are set
on interface <ethn>. See ifconfig(8).
```

IPv4

Only IPv4 is supported. If another address family is specified for the network, PanPulse will report an error such as the following:

```
Interface <ethn> has addrtype AF_INET6, skipping
```

Separate Networks

Each network interface card must be on a separate network. If this is not the case, PanPulse will report the following error:

```
Interface <ethn> will not be used by panpulse because it
exists on the same network as interface <ethn>. This
configuration is not supported by HP StorageWorks Clustered
File System.
```

Port 8940

Only one instance of PanPulse can be running on port 8940 on a server. If another application is using that port or another instance of PanPulse is started, the following error will be reported.

```
Unable to bind on port 8940. Please make sure that this is
the only copy of panpulse running on this server.
```

mxinit Messages

mxinit prints a series of messages when it cannot complete an operation.

For example, **mxinit** must unmount all PSFS filesystems before it shuts down HP Clustered File System. If a filesystem is busy, it cannot be unmounted. **mxinit** will then issue a message about the filesystem, followed by additional messages stating that **mxinit** could not complete the various steps involved in shutting down HP Clustered File System.

Similarly, if **mxinit** cannot start HP Clustered File System because, for example, the system is out of resources, you will see a series of messages reporting that start-up actions could not be completed.

Index